

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 19, 2005, 00:57:16 ; Search time 194 Seconds
(without alignment)
3188.617 Million cell updates/sec

Title: US-09-889-325-4
Perfect score: 6424
Sequence: 1 MERLRDVRERLQAMERAFRR.....HLSPHALVGLATEELLQVAR 1208

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 1612378 seqs, 512079187 residues
Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt Q3:*
1: uniprot_aprot:*
2: uniprot_crembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6424	100.0	1208	1 RCQ4 HUMAN	094761 homo sapien
2	3988	62.1	1216	2 Q75NR7	075nr7 mus musculu
3	3944	61.4	1216	2 Q99PV9	099pv9 mus musculu
4	3923	61.1	741	2 Q96F55	096f55 homo sapien
5	3469	54.0	652	2 Q96DW2	096dw2 homo sapien
6	3439	53.5	1151	2 Q76MT1	076mt1 mus musculu
7	1723.5	26.8	1579	2 Q9VSE6	09vse6 drosophila
8	1713.5	26.7	1552	2 Q7PMM8	07pmm8 anopheles g
9	1680.5	26.2	1530	2 Q9NH11	09nh11 drosophila
10	1172	18.2	340	2 Q7YR85	07yr85 bos taurus
11	971.5	15.1	927	2 Q7X829	07x829 oryza sativ
12	897.5	14.0	870	2 Q8W028	08w028 arabidopsis
13	872	13.6	941	2 Q9C6N0	09c6n0 arabidopsis
14	772	12.0	899	2 Q84RP1	084rp1 oryza sativ
15	761	11.8	874	2 Q7FAD0	07fad0 oryza sativ
16	635	9.9	704	2 Q7MID7	07mid7 vibrio vuln
17	631.5	9.8	604	2 Q83IW3	083iw3 shigella fl
18	631.5	9.8	608	1 RECO ECOGL	07u49 escherichia
19	631.5	9.8	611	2 Q7UB49	08fme escherichia
20	631.5	9.8	611	2 Q8FBM6	08xni escherichia
21	631.5	9.7	611	2 Q8X8N1	08db16 vibrio vuln
22	626	9.7	639	2 Q8BDB16	08czh9 erwinia car
23	617.5	9.6	654	2 Q87MB5	087mb5 pseudomonas
24	613.5	9.6	608	2 Q8CZH9	0823b2 salmonella
25	613	9.5	642	2 Q87U96	087u96 salmonella
26	609.5	9.5	609	2 Q823B2	0823b2 salmonella
27	607.5	9.5	608	1 RECO SALTU	0823b2 salmonella
28	604	9.4	654	2 Q8BEK1	08beek1 shewanella
29	599	9.3	746	2 Q7NH48	07nh48 gloeobacter
30	593	9.2	610	2 Q66FY3	066fy3 yersinia ps
31	593	9.2	610	2 Q8ZAG8	08zag8 yersinia ps

32	586.5	9.1	608	2 Q7WY05	07wy05 photorhabdu
33	586	9.1	602	2 Q8BY35	08by35 lactobacill
34	584.5	9.1	641	2 Q6LXK0	06lxk0 photobacter
35	581.5	9.1	620	2 Q9KVF0	09kvf0 vibrio chol
36	580.5	9.0	749	2 Q7UR05	07ur05 rhodospirill
37	578.5	9.0	632	1 RECO PASMU	09c121 pasteurella
38	578	9.0	619	1 RECO HABIN	071359 haemophilus
39	573.5	8.9	615	2 Q6LLP9	06llp9 photobacter
40	573	8.9	637	2 Q6SQS2	06sq2 mannheimia
41	571.5	8.9	1002	2 Q7UEX6	07uex6 geobacter s
42	570	8.9	603	2 Q74ER2	074er2 geobacter s
43	566.5	8.8	611	2 Q87KH0	087kh0 vibrio para
44	560.5	8.7	601	2 Q7VNM7	07vnm7 haemophilus
45	560.5	8.7	602	2 Q8UJ77	08uj77 agrobacteri

ALIGNMENTS

RESULT 1
RCQ4_HUMAN STANDARD; PRT; 1208 AA.
ID RCQ4_HUMAN
AC 094761;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE ATP-dependent DNA helicase Q4 (RecQ protein-like 4) (RecQ4).
GN Name=RECQ4; Synonyms=RECQ4;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid:9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=99097344; PubMed=9878247; DOI=10.1006/geno.1998.5595;
RA Kitao S., Ohnagi I., Ichikawa K., Goto M., Furuchi Y., Shimamoto A.;
RT "Cloning of two new human helicase genes of the RecQ family:
RT biological significance of multiple species in higher eukaryotes.";
RL Genomics 54:443-452(1998).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=20021764; PubMed=10552928; DOI=10.1006/geno.1999.5599;
RA Kitao S., Lindor N.M., Shiratori M., Furuchi Y., Shimamoto A.;
RT "Rothmund-Thomson syndrome responsible gene, RECQ4: genomic structure
RT and products.";
RL Genomics 61:268-276(1999).
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- DISEASE: Defects in RECQ4 are a cause of Rothmund-Thomson
CC syndrome (RTS) [MIM:268400]. A disease characterized by
CC dermatological features such as atrophy, pigmentation, and
CC telangiectasia and frequently accompanied by juvenile cataract,
CC saddle nose, congenital bone defects, disturbances of hair growth,
CC and hypogonadism.
CC -!- SIMILARITY: Belongs to the helicase family. RecQ subfamily.
CC -!- DATABAS: NMB=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC WWW="http://www.intbioogen.fr/services/chromcancer/Genes/RECQ4ID285.html".
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; AB006532; BAA74453.1; -;
DR EMBL; AB025545; BAA68899.1; -;
DR Genew; HGNC:9949; RECQ4.
DR MIM; 603780; -;
DR WU; 268400; -;
DR GO; GO:0003678; F:DNA helicase activity; TAS.
DR GO; GO:0007275; P:development; TAS.

DR GO; GO:0006281; P.DNA repair; TAS.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002464; DEAH box.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR004589; RecQ.
 DR Pfam; PF00270; DEAD; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR TIGRFAMs; TIGR00614; recQ fam; 1.
 DR PROSITE; PS00690; DEAH ATP HELICASE; FALSE NEG.
 KM ATP-binding; Cataract; Helicase; Hydrolyase; Nuclear protein.
 FT NP_BIND 502 509 ATP (Potential).
 FT SITE 605 608 DEAH box.
 SQ SEQUENCE 1208 AA; 133076 MW; CB809A7765AB48A1 CRC64;

Query Match 100.0%; Score 6424; DB 1; Length 1208;
 Best Local Similarity 100.0%; Pred. No. 1.4e-297;
 Matches 1208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MERLDVRERLQAMERAFRRQRGRPSDDVEAARETRALYREYRLKRTTGOAGGGLR 60
 DB 1 MERLDVRERLQAMERAFRRQRGRPSDDVEAARETRALYREYRLKRTTGOAGGGLR 60

QY 61 SSESIPAAAEBAEPERCWPHLNRAATKSPPTPGSRGSGSPDYQRLKANLKTGLQNG 120
 DB 61 SSESIPAAAEBAEPERCWPHLNRAATKSPPTPGSRGSGSPDYQRLKANLKTGLQNG 120

QY 121 PALGRPWFPLGRASSKASTPKPGTGPVSPFAEKVSDDEPPOLPEPQPRGRLQHLQASIS 180
 DB 121 PALGRPWFPLGRASSKASTPKPGTGPVSPFAEKVSDDEPPOLPEPQPRGRLQHLQASIS 180

QY 181 QRLGSLDPGWLQRCHEVDPFLGAPKACRPDLGSEESQLLIPGESAVLGPAGSGQPEAS 240
 DB 181 QRLGSLDPGWLQRCHEVDPFLGAPKACRPDLGSEESQLLIPGESAVLGPAGSGQPEAS 240

QY 241 AFOEVSIRVGSQPPSSGGEKRWMBEPSPQVOOESSQAGPSEEGANVAVEDPPG 300
 DB 241 AFOEVSIRVGSQPPSSGGEKRWMBEPSPQVOOESSQAGPSEEGANVAVEDPPG 300

QY 301 EPVQAPQPPQSSPNRYHGLSPSSQARAGKAGTAAPHITPRLLRHNGVYRLNMKG 360
 DB 301 EPVQAPQPPQSSPNRYHGLSPSSQARAGKAGTAAPHITPRLLRHNGVYRLNMKG 360

QY 361 KHYVGRALRSRLLRKQWKQKWKKEGFCGGGATVTTKESCFINEQFDHMAAQCPRA 420
 DB 361 KHYVGRALRSRLLRKQWKQKWKKEGFCGGGATVTTKESCFINEQFDHMAAQCPRA 420

QY 421 SEEDTDVGPPEPLVPSPQVPEPSLDPTVLPYLSGSPGQLAETPAEVFQALEQGHQ 480
 DB 421 SEEDTDVGPPEPLVPSPQVPEPSLDPTVLPYLSGSPGQLAETPAEVFQALEQGHQ 480

QY 481 FRPGQERAWRIISGISTLLVLPYTGAGKSLCYQLPALYRSRSPCLTLVSPPLSLMDQ 540
 DB 481 FRPGQERAWRIISGISTLLVLPYTGAGKSLCYQLPALYRSRSPCLTLVSPPLSLMDQ 540

QY 541 VSGLPPLCLACIHSGMTKQKQRESVLQKIRAAQVHVLMTPEALVAGAGLPPAAQLPPVA 600
 DB 541 VSGLPPLCLACIHSGMTKQKQRESVLQKIRAAQVHVLMTPEALVAGAGLPPAAQLPPVA 600

QY 601 FACIDAHCLSQMSNHPRCYLAVCVLAERMGVHFLGTLATATRTSDVAQHLAAVE 660
 DB 601 FACIDAHCLSQMSNHPRCYLAVCVLAERMGVHFLGTLATATRTSDVAQHLAAVE 660

QY 661 EPLDHPAPVPTNLHLSVSMRDTQDALTLTLOGKSFQNLDSITIIYCNREDTERLAALL 720
 DB 661 EPLDHPAPVPTNLHLSVSMRDTQDALTLTLOGKSFQNLDSITIIYCNREDTERLAALL 720

QY 721 RTCLHAAMVPGSGGAPKTTAAEYHAGMSRERRVQRAFMOQLRVVAATVAFGKGLR 780
 DB 721 RTCLHAAMVPGSGGAPKTTAAEYHAGMSRERRVQRAFMOQLRVVAATVAFGKGLR 780

QY 781 PDVRAVLHLGLPPSPFSYVOAVGRAGRDGPALHCHLFLQPGQDLRELRHVADSTDFL 840
 DB 781 PDVRAVLHLGLPPSPFSYVOAVGRAGRDGPALHCHLFLQPGQDLRELRHVADSTDFL 840

QY 841 AVKRLVQVFPACTCTCTRPSPBOGAVGERPVKYPPOEABQLSHQAAPRREVCNGH 900
 DB 841 AVKRLVQVFPACTCTCTRPSPBOGAVGERPVKYPPOEABQLSHQAAPRREVCNGH 900

QY 901 ERALPIQLTVQALMDPEBAITLLCYLELHPHHMIELATTYTHRLNCPGSPALQALA 960
 DB 901 ERALPIQLTVQALMDPEBAITLLCYLELHPHHMIELATTYTHRLNCPGSPALQALA 960

QY 961 HRCPLAVCLAQQLPEBDGQSSSVFPMVKLVDSMGWELASVRBALCOLQMDHPPRGV 1020
 DB 961 HRCPLAVCLAQQLPEBDGQSSSVFPMVKLVDSMGWELASVRBALCOLQMDHPPRGV 1020

QY 1021 RRGTVLVEFSELAFLHLSPGDLTAEEKQICDPLYGAVQAEQALRLRRTFOAFHSV 1080
 DB 1021 RRGTVLVEFSELAFLHLSPGDLTAEEKQICDPLYGAVQAEQALRLRRTFOAFHSV 1080

QY 1081 AFPSGCPCLQODDERSTRLDLGRFYEEBEGQPGMEDAQGPQARLQMDWQDV 1140
 DB 1081 AFPSGCPCLQODDERSTRLDLGRFYEEBEGQPGMEDAQGPQARLQMDWQDV 1140

QY 1141 CDIRQFLSLRPEKFSRAVARIFHGSPCYPAOVYQODRRFMRKYHLHSFHALVGLAT 1200
 DB 1141 CDIRQFLSLRPEKFSRAVARIFHGSPCYPAOVYQODRRFMRKYHLHSFHALVGLAT 1200

QY 1201 BELLOVAR 1208
 DB 1201 BELLOVAR 1208

RESULT 2
 075NR7
 ID 075NR7 PRELIMINARY; PRT; 1216 AA.
 AC 075NR7;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DE RecQ helicase protein-like 4.
 GN Name=Recq14;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ohata T., Fukumura R., Araki R., Abe M.;
 RL Submitted (Mar-2004) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AB175741; BAI4289.1; .
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
 DR GO; GO:0016787; F:hydrolyase activity; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR GO; GO:0006259; P:DNA metabolism; IEA.
 DR InterPro; IPR002086; Aldehyd_dehydrog.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR011545; DEAD/DEAH_N.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR004589; RecQ.
 DR InterPro; IPR001424; SOD_CU_ZN.
 DR InterPro; IPR001878; Znf_CCHC.
 DR Pfam; PF00270; DEAD; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR Pfam; PF00098; zf_CCHC; 1.
 DR PRINTS; PR00939; C2HCZFNINGER.
 DR SMART; SM00487; DEXDC; 1.
 DR SMART; SM00490; HELIC; 1.
 DR SMART; SM00343; Znf_C2HC; 1.
 DR TIGRFAMs; TIGR00614; recQ fam; 1.
 DR PROSITE; PS00070; ALDEHYDE DEHYDR CYG; UNKNOWN_1.
 DR PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
 DR PROSITE; PS00158; ZF_CCHC; 1.
 KM ATP-binding; Helicase; Hydrolyase.
 SQ SEQUENCE 1216 AA; 135081 MW; 0A7336EBAE167306 CRC64;

Query Match 62.1%; Score 3988; DB 2; Length 1216;
 Best Local Similarity 64.4%; Pred. No. 1.1e-181;
 Matches 801; Conservative 112; Mismatches 259; Indels 72; Gaps 13;

1 MERLADVBERLOQWERAERARORRRPODDVEAAEETRALYREYRTLKRTTGAAGGIR 60
 1 MERLATVABARLOEWERARARLRGRPAKDVAAEETRALYREYRNKQAVROADDNR 60
 61 SSE-SLPAABEAPRRCWGHPLNRAATKSPQTFGRSRQSVVPDYGORLANKLGTLOA 119
 61 VLEGLAAABEAPRRCWGHPLNRAATKSPQTFGRSRQSVVPDYGORLANKLGTLOA 120
 120 GPALGRWRPPLGRASASTPRPCTGPVPSAEKVSDEPPOLPFPQPRPGLQLOASTL 179
 121 GPTOSRKLQLOKRSISTVPAPRPPGSKTESPCPDADDLPRVPRPRPLGLOOLRSSL 180
 180 SORLGLDPGWLQRCHEVPDFLQAPKACRPDLGSEESQLIPGESAVLGPAGS-----Q 235
 181 SRRLTSLDPGMERCHNRVSDLEVPAGCGLDLSAEESPQMSGKYNLADPDIOGEVSQ 240
 236 GPESAFGEVSIRVSPQSSSGEKRRWNEEPWESPAPVOQESSQAGPSSGAGAVAVE 295
 241 SPEALAQPAQLSQSPKSKNGRRKRKNEK-GEFDAQ-DQPSGAGPLSGABATVHG 298
 296 EDPPEEPVQAPRQPCSSPSNRHYGLSSQABAGKAGTAPLIFPLAHNDGNVYR 355
 299 QDPPEEPVQAVNPQPCNS-----SNQARTKAGTTHLHASPPLADRGNYIR 347
 356 LNMKQHYVGRGALSRLLRKQAMKMKKECEGSGGATVTTESCPFLNEQDFHMAAQ 415
 348 LNMKRRVTVRVANGRLLRKQWQKMKKQAAFGSGSPRATDIDTCRCQFGHMASQ 407
 416 CRRPA-----SEEDTAVP-----EPVSPQVPEVBSL 446
 408 CSQPGPTLVQEGGRDQKQISTLEBAVQRTGTASCHHSGETQPAABELQ-VRHCPT- 465
 447 DPTVPLVSLSGGLAETRAVFAQALBGLHQAAPRQGERAVMVLISITLVLPTCA 506
 466 --PMSPLPFGGLGVAETPAVFALELRLGTRARPCGERIMLILSITLVLPTCA 523
 507 GKSLCYQPLALVYSRRSPCLTLVPSPLSLMDQVSGLPCLKAACIHSGMTKROESYL 566
 524 GKSLCYQPLALVYSRRSPCLTLVPSPLSLMDQVSDLPCLKAACIHSGMTKROESYL 583
 567 QKIRAAQVHMLTPEALVGA-----GGLPRAQQLPVAACIDEAHCLSONSHNFRPCY 621
 584 KKVRAAQVHMLTPEALVGA-----GGLPRAQQLPVAACIDEAHCLSONSHNFRPCY 643
 622 LRVCVLBERMGVHCFGLTATATRTASDVQHLAVAEPRDLHGPAVPTLHLISVMD 681
 644 LRVCVLBERMGVHCFGLTATATRTASDVQHLAVAEPRDLHGPAVPTLHLISVMD 703
 682 RDTDOALTLTGKGFONLDSIIICYNRREDTERIAALLRTCLHAAWVSGSGRAPKTA 741
 704 RDSDALVTLLGDFRITLDSVIITCRKQTERVAAALITLISWGSRRPGCGEBAIA 763
 742 EAYHAGMSRRRRVQARAFMOQLRVVAVTAFMGCLDRPDVRAVLHLGLPPSFESYVOA 801
 764 EAYHAGMSRRRRVQARAFMOQLRVVAVTAFMGCLDRPDVRAVLHLGLPPSFESYVOA 823
 802 VGRARDDQAPACHFLQPOGEBDLAELRHVAVADSTPLAVKLVQVPRPATCTCTRRP 861
 824 IGRARDDQAPACHFLQPOGEBDLAELRHVAVADSTPLAVKLVQVPRPATCTCTRRP 878
 882 SEQEGVAGGERVPKYPQEAQLS-HQAPGRPRVCMGHEALPLQLTVQALDMPREAI 920
 879 -----GRVSSSPFEVEKESGQTYVPLGQACLGHBAALPVQSTVQALDMPREAI 929
 921 ETLICYLBIHPRHMLLELATYTHCLNCPGAPQLOALAHRCPLAVCLAQOLBEDPQO 980
 930 ETLICYLBIHPRHMLLELATYTHCLNCPGAPQLOALAHRCPLAVCLAQOLBEDPQO 989

Query Match 61.4%; Score 3944; DB 2; Length 1216;
 Best Local Similarity 63.9%; Pred. No. 1.4e-179;
 Matches 795; Conservative 110; Mismatches 267; Indels 72; Gaps 13;

981 GSSVFPDMVKLVDSMGWELASVRRALCOLQMDHEPRTGVRGTVLVESELAFTLRSP 1040
 990 GRSSIEFVEVLADSMGKSLASVRQALHQLKMDPEPKGAAGCTGLVLFESLAFHLNR 1049
 1041 GDLTAEKDOICDPLVGVQARERQALRLRRTQAFHSVAFPSGCPCLCEQDEERSTRL 1100
 1050 GDLTAEKDOICDPLVGVQARERQALRLRRTQAFHSVAFPSGCPCLCEQDEERSTRL 1109
 1101 KDLAGRYEEBEGPFGMGEDAGCPGQALQDMEDVRCDIRGFLSLRPREKSSRAV 1160
 1110 KTLVSYFEEBEE-EETETMTDQGPQOTLOQMEDDIRDVRQLSLRPERFSRAV 1168
 1161 ARIFGIGSPCYPAOVYGODRFRMKYHLHSPHALVATSEEL 1204
 1169 ARIFGIGSPCYPAOVYGODRFRMKYHLHSPHALVATSEEL 1212

RESULT 3
 ID 099PV9 PRELIMINARY; PRT; 1216 AA.
 AC 099PV9;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DS RecQ helicase protein-like 4.
 GN Name=Recq4; Synonyms=RECQ4;
 OS Mus musculus (Mouse);
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Abe M., Ohnata T.;
 RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB042528; BAB32696.1; JOINED.
 DR HSSP; P15043; 10YY.
 DR MGD; MGI:1931028; Recq4.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
 DR GO; GO:0016787; F:hydrolyase activity; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR GO; GO:0006259; P:DNA metabolism; IEA.
 DR InterPro; IPR02086; Aldehyd_dehydrog.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR011545; DEAD/DEAH_N.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR004589; RecQ.
 DR InterPro; IPR001424; SOD CU ZN.
 DR InterPro; IPR001878; Znf_CCHC.
 DR Pfam; PF00270; DEAD; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR Pfam; PF00098; ZF-CCHC; 1.
 DR PRINTS; PR00939; C2HCZNFINGER.
 DR SMART; SM00487; DEXDC; 1.
 DR SMART; SM00490; HELIC; 1.
 DR SMART; SM00343; ZNF_C2HC; 1.
 DR TIGRfam; TIGR00614; recQ_fam; 1.
 DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
 DR PROSITE; PS00087; SOD CU ZN 1; UNKNOWN_1.
 DR PROSITE; PS0158; ZF_CCHC; 1.
 KW ATP-binding; Helicase; Hydrolyase.
 SQ SEQUENCE 1216 AA; 135122 MW; 431D79843BAEC2AF CRC64;

Query Match 61.4%; Score 3944; DB 2; Length 1216;
 Best Local Similarity 63.9%; Pred. No. 1.4e-179;
 Matches 795; Conservative 110; Mismatches 267; Indels 72; Gaps 13;

1 MERLADVBERLOQWERAERARORRRPODDVEAAEETRALYREYRTLKRTTGAAGGIR 60
 1 MERLATVABARLOEWERARARLRGRPAKDVAAEETRALYREYRNKQAVROADDNR 60

```

Db      61  VLEQGLAAAEBAQSPSCWPHLSRAATQNTQSMWKSQSLSSVQCYGKRLKANLKNITQT 120
Qy      120  GPALGRRRWPLGRASASTPKPRGTPVPSPAEKVSDEPDPLEPPORPPGLQLOAST 179
Db      121  GPTQGRKIQLOQRSLSTVPAPRPSPKTESPCPDADDLAPVPBPRLPGLOQLRSSL 180
Qy      180  SQRLSLDPGWLQRCHESEVPDFLGAPKACRPDLGSEESQLLPGESAVLGPAGS----Q 235
Db      181  SRRLSLDPGWLERCHNRVSDLEVPGACGLDSABESQPMQSGKVNADPDIQSEVSNQ 240
Qy      236  GPEASAFQEVSTIRVSPQSSSGEKKRWNEEPWESPAPVOQESSQAGPSSGAGAVAYE 295
Db      241  SPEALAQCPAQVLSQSPKSKINSKGRKRKNK-KGDFAQ-DQPSGAGPISGAGATVYG 298
Qy      296  EDPPEEPVQAOPOPCCSSSNRNHGLSSQABAKCKEKPRLHIFPLAHNDGNVYR 355
Db      299  QDPPEEPVQVNPQPCNS-----SNQARTKAKGTHHLASPPRPSALDGNVYR 347
Qy      356  LNMKGKHYVRGRALSSRLRKQAMKQMKKGECEGSGGATVTTKESCPLNQPFHMAAQ 415
Db      348  LNMKGRKRFVRVANGRLLRKQWKQKMKKQAAFGSGSPRATDVTDCRCQPFHMASQ 407
Qy      416  CPRPA-----SEEDTAVGP-----EPLVSPQPVPEVPSL 446
Db      408  CSQPEPTLVQEGGRDDKQPISTLEEVNQRTGTASCHNSGEGTQPAAPBLQ-VPHCPT- 465
Qy      447  DPTVPLVSLSGSGLAETPAEVPALBQAGAPRQGERVNMVILISITLVLPYGA 506
Db      466  --PMSBPLPGPLGQVLETPAEVPQALELGRARPPGERKIMVILISITLVLPYGA 523
Qy      507  GKSICYOLPALLYSRRSPCLTVVSPPLSLMDQVSGPRLCAACIHSGMTKRORESVL 566
Db      524  GKSICYOLPALLIYAGRSPLCLTVVSPPLSLMDQVSDLPSCIAKACIHSGMTKRORESVL 583
Qy      567  QKIRAAQVHTMLTPBALVGA-----GGLPRAQLRPVAPACIDEAHCSIGMSHNFRPY 621
Db      584  KKVRAQVHTLVLSPEALVGCARPGSGSLPQAOLPPIAFACIDEVHCISQMSHNFRPY 643
Qy      622  LRVCVLPRMGVNHQELGTATARTRTASDVNQHVAVNEERPLHGPAPPTLHLSVSD 681
Db      644  LRVCVLPRMGVNHQELGTATARTRTASDVNQHVAVNEERPLHGPAPPTLHLSVSD 703
Qy      682  RPTDQALTLTLOGKFFONDSIIYCNRRDTERIALARTGLAHAMVWGSGRAPKTTA 741
Db      704  RPSDQALVTLTLOGDRRTIDSIIYCTRBRIGMGLALRTGLSNWGSRRPGCGEALTA 763
Qy      742  EAYHAGMSRRRRVQARFMQGLRVVATVAFMGGLDRPDVAVLHGLPPSFESYVQA 801
Db      764  EAYHAGMSQERRRRVQAFMRGHLRMVAVTVAFGMGLDRPDVAVLHGLPPSFESYVQA 823
Qy      802  VVARARDDQRAHCHLFLQFQEGEDLRLRHNHVNADSDPLAVRIVQVRPACSTCTRRP 861
Db      824  IRRAGDGRRAHCHLPMRQGBELRLRHNHVNADSDPLAVRIVQVRPACSTCTCTRRP 878
Qy      862  SEQEGAVGGERVVPKTPQOEAOLS-HQAPRPARYVCMGHEBALPQLTVQALDMEBEAL 920
Db      879  -----GRVSKSSPEVEKENSQOQYUPLVGAQCSGHEBALPVQSVQALDMEBEAL 929
Qy      921  ETLTCLYELHPRNHWELELATTTTHCRKLRGGRPAQLOALAHRCRPLAVGLAQOLREDDQ 980
Db      930  ETLTCLYELHPRNHWELELPMTYAQCSYHCLGSAQLOALAHRCRPLAACQAKMPKDTISQ 989
Qy      981  GSSSVFEDVVKLVDSMGWELASVRYRLCOLMDHNERGTGVRGTVLVEFSELAFLHSP 1040
Db      990  GSSSVFEGVVELADSGWKLASVRYRLHQLKMDPEKKAAAGQTVLVFSELAFLHSP 1049
Qy      1041  GDLTAEEKQICDPLTGVQADAEQALARTFTQAFHSVAPSPSCPCLEQDDEERSTL 1100
Db      1050  GDLTBEKQICDPLTGNVQAREHKLALHOMSKAFRVAFPSCPCLEQDDEERSTL 1109
Qy      1101  KULLGRYFEESQGEFGMEDAAGPFGAQLQDWEQVRCIRQPLSLRPEKFSRRAY 1160

```

```

Db      1110  KTVSYVEEER-ESEETMTDQGPQGTQLQDWEQIRDVNQLSLRPEERSGRAV 1168
Qy      1161  ARIFHGSPCYPAQVYQGDPRFMRKYHLHLSFHALVGLATEELL 1204
Db      1169  ARIFHGSPCYPAQVYGLDRFRMRKYHLHLSFHALVGLATEELL 1212

RESULT 4
ID 096F55 PRELIMINARY; PRT; 741 AA.
AC 096F55;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE RECOL4 protein (Fragment).
GN Name=RECOL4;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RK MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strusberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schenker C.F., Bat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marnett K., Farmer A.A., Rubin G.M., Hong L.,
RA Staphleon M., Soares M.B., Bonaldo M.F., Caavaat T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carantini P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Faney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Boultard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buterfield Y.S.,
RA Krzyzanski M.I., Skalski U., Smallus D.E., Scherch A., Schein J.B.,
RA Jones S.J., Maira M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strusberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC011602; AAH11602.2; -.
DR HSSP; P15043; IOY.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0016787; F:hydrolyase activity; IEA.
DR GO; GO:0006259; F:DNA metabolism; IEA.
DR GO; GO:0006259; F:DNA metabolism; IEA.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR001545; DEAD/DEAH_N.
DR InterPro; IPR001545; DEAD/DEAH_N.
DR InterPro; IPR001545; DEAD/DEAH_N.
DR pfam; PF00270; DEAD_1.
DR pfam; PF00271; Helicase_C_1.
DR SMART; SM00487; DEXDC_1.
DR SMART; SM00490; HELIC_C_1.
DR TIGRfam; TIGR00614; recQ_fam_1.
KW ATP-binding; Helicase; Hydrolase.
FT NON TER
SQ SEQUENCE 741 AA; 82502 MW; 001E5EA21E80B6A7 CRC64;

Query Match 61.1%; Score 3923; DB 2; Length 741;
Best Local Similarity 100.0%; Pred. No. 8,1e-179; Indels 0; Gaps 0;
Matches 741; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 468 EVFOALAEQHOAFRPGGERAVNRILSGISTLLVPTGAGSKLCYOLPALIYSRSPCLT 527
Db 1 EVFOALAEQHOAFRPGGERAVNRILSGISTLLVPTGAGSKLCYOLPALIYSRSPCLT 60
QY 528 LVSPSLIMDDQVSGLPCLKAACIHSGMTRKQRESVLOKIRAAOVHVLMTPEALVGA 587
Db 61 LVSPSLIMDDQVSGLPCLKAACIHSGMTRKQRESVLOKIRAAOVHVLMTPEALVGA 120
QY 588 GGLPPRAOOLPVAFACIDBAHCLSGMSNFRPCYLAVCVLAERMGVHFLGTTATATRR 647
Db 121 GGLPPRAOOLPVAFACIDBAHCLSGMSNFRPCYLAVCVLAERMGVHFLGTTATATRR 180
QY 648 TASDVAQHIAVAEPEPLHGPAPVPTLHLSVSMNDRTDQALLTLQKGFONLDSITTC 707
Db 181 TASDVAQHIAVAEPEPLHGPAPVPTLHLSVSMNDRTDQALLTLQKGFONLDSITTC 240
QY 708 NREDETERIALLRCTLAHAAMPVSGSGRAPKTTAEAYHAGMSRERRRVOGAFMOQLAV 767
Db 241 NREDETERIALLRCTLAHAAMPVSGSGRAPKTTAEAYHAGMSRERRRVOGAFMOQLAV 300
QY 768 VVATVAFGGLDRPDVRAVLHLGLPSPFSYVOAVRAGRDGPANCHLFLQPGEDLRE 827
Db 301 VVATVAFGGLDRPDVRAVLHLGLPSPFSYVOAVRAGRDGPANCHLFLQPGEDLRE 360
QY 828 LRRHYADSTDFLAVKRLVQVVPACTCTPRPSQEGAVGGERPVKYPQDEAQLSH 887
Db 361 LRRHYADSTDFLAVKRLVQVVPACTCTPRPSQEGAVGGERPVKYPQDEAQLSH 420
QY 888 QAPRRVRCMGERALPIQLTVQALDMEBEAIELLCVLELPHHMLATTTTCHL 947
Db 421 QAPRRVRCMGERALPIQLTVQALDMEBEAIELLCVLELPHHMLATTTTCHL 480
QY 948 NCPGPAQLOALAHRCPLAVCLAOQLPEDPGQSSSVPRDYKVLVDSMGWELASVRL 1007
Db 481 NCPGPAQLOALAHRCPLAVCLAOQLPEDPGQSSSVPRDYKVLVDSMGWELASVRL 540
QY 1008 COLQMDHEPRTGVRGTGLVPSFLAFLHRSKGLTAEKQICDPLGRVQARRQAL 1067
Db 541 COLQMDHEPRTGVRGTGLVPSFLAFLHRSKGLTAEKQICDPLGRVQARRQAL 600
QY 1068 ARLRRTFOAFHSAFSPGCPCLCEQODEERSTRKDLGRYFEEBEGQEPGMEDAGPBP 1127
Db 601 ARLRRTFOAFHSAFSPGCPCLCEQODEERSTRKDLGRYFEEBEGQEPGMEDAGPBP 660
QY 1128 GQARLDQMEDQVRCIRQFLSLRPEBKFSRAVARIFHGIGSPCYAUVGYDDRRFWRY 1187
Db 661 GQARLDQMEDQVRCIRQFLSLRPEBKFSRAVARIFHGIGSPCYAUVGYDDRRFWRY 720
QY 1188 LHLSPHALVGLATELLOVAR 1208
Db 721 LHLSPHALVGLATELLOVAR 741

```

```

RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Bromslein M.J., Usdin T.B., Toshiyuki S., Carantini P., Prange C.,
RA Rana S.B., Loguettano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.V., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huliyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kertman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013277; AAH13277.2; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; Helicase_C; 1.
DR SMART; SM00490; HELIC_C; 1.
SQ SEQUENCE 652 AA; 72987 MW; 328B784BD39735F7 CRC64;

Query Match 54.0%; Score 3469; DB 2; Length 652;
Best Local Similarity 100.0%; Pred. No. 2.8e-157;
Matches 652; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 557 MRRKRESVLOKIRAAOVHVLMTPEALVGAAGLPPAOLPVAFACIDBAHCLSGMSN 616
Db 1 MRRKRESVLOKIRAAOVHVLMTPEALVGAAGLPPAOLPVAFACIDBAHCLSGMSN 60
QY 617 FRPCYLAVCVLAERMGVHFLGTTATATRRASDVAQHIAVAEPEPLHGPAPVPTNL 676
Db 61 FRPCYLAVCVLAERMGVHFLGTTATATRRASDVAQHIAVAEPEPLHGPAPVPTNL 120
QY 677 SVSMNDRTDQALLTLQKGFONLDSITTCNREDETERIALLRCTLAHAAMPVSGSGRA 736
Db 121 SVSMNDRTDQALLTLQKGFONLDSITTCNREDETERIALLRCTLAHAAMPVSGSGRA 180
QY 737 PRTAEAYHAGMSRERRRVOGAFMOQLRVVAVVAFGMLDRPDVRAVLHLGLPSPFE 796
Db 181 PRTAEAYHAGMSRERRRVOGAFMOQLRVVAVVAFGMLDRPDVRAVLHLGLPSPFE 240
QY 797 SYVOAVGRAGRDGPANCHLFLQPGEDLRELRHVHADSTDFLAVKLVQVFPACTCT 856
Db 241 SYVOAVGRAGRDGPANCHLFLQPGEDLRELRHVHADSTDFLAVKLVQVFPACTCT 300
QY 857 CTRPSQEGAVGGERPVKYPQDEAQLSHQAPRRVRCMGERALPIQLTVQALDMP 916
Db 301 CTRPSQEGAVGGERPVKYPQDEAQLSHQAPRRVRCMGERALPIQLTVQALDMP 360
QY 917 BEAIELLCVLELPHHMLATTTTCHLRCNCPGPAQLOALAHRCPLAVCLAOQLPE 976
Db 361 BEAIELLCVLELPHHMLATTTTCHLRCNCPGPAQLOALAHRCPLAVCLAOQLPE 420
QY 977 DPGQSSSVPRDYKVLVDSMGWELASVRLAQLOMDHEPRTGVRGTGLVPSFLAFL 1036
Db 421 DPGQSSSVPRDYKVLVDSMGWELASVRLAQLOMDHEPRTGVRGTGLVPSFLAFL 480
QY 1037 LRSPEDLTAEKQICDPLGRVQARRQALRLRRTFOAFHSAFSPGCPCLCEQODEER 1096
Db 481 LRSPEDLTAEKQICDPLGRVQARRQALRLRRTFOAFHSAFSPGCPCLCEQODEER 540
QY 1097 STRLMDLGRYFEEBEGQEPGMEDAGPBPQALQMDQVRCIRQFLSLRPEBKFS 1156

```

|||||
541 STRLKDILGRYFEEBEGGEPGGMEDAQGPBGQARLDQWEDQVRCIRQFLSLRPEKPS 600
DB
QY 1157 SAAVARIFHGIGSPCPYPAOVYODRRFMRKYLHLSTHALVGLATELLOVAR 1208
DB 601 SAAVARIFHGIGSPCPYPAOVYODRRFMRKYLHLSTHALVGLATELLOVAR 652
RESULT 6
076MT1 PRELIMINARY; PRT; 1151 AA.
ID 076MT1
AC 076MT1
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DE RecQ4 homolog (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Abe M., Ohnata T.,
RL Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB039882; BAD1131.1; -
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0008026; P:ATP-dependent helicase activity; IEA.
DR GO; GO:0016787; F:hydrolyase activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006259; P:DNA metabolism; IEA.
DR InterPro; IPR002086; Aldehyd_dehydroy.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004589; RecQ.
DR InterPro; IPR001424; SOD_CU_ZN.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00270; DEAD; 1. Znf_CCHC.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00098; zf-CCHC; 1.
DR PRINTS; PR00939; C2HCNFINGER.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
DR SMART; SM00343; Znf_C2HC; 1.
DR TIGRFAMs; TIGR00614; recQ_fam; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYD_CYS; UNKNOWN_1.
DR PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
DR PROSITE; PS00158; ZF_CCHC; 1.
KW ATP-binding; Helicase; Hydrolase.
FT NON TER 1
FT 1151 1151
SQ SEQUENCE 1151 AA; 126528 MW; 97CF481ABC4B9A CRC64;
Query Match 53.5%; Score 3439; DB 2; Length 1151;
Best Local Similarity 59.9%; Pred. No. 1.4e-155;
Matches 707; Conservative 103; Mismatches 248; Indels 122; Gaps 14;
QY 16 RAFFRRGRGRRPSQDDVEAARPEETRALYREYRTLKRTTGAAGGLRSE-SLPAAAEARE 74
DB 1 RAFFRRGRGRRPAQDEAARPEETRALYREYRTLKQAVRADDRHRLLEQSLAEAAEAE 60
QY 75 PRCMGHILNRAATKSPPTPGRSRQGSVPDYQRLKANTLKGTLQAGPALGRPRPLGRAS 134
DB 61 PCMGHILNRAATQNTQSMFKQSLSSVQDYGRKLTANLKNTTQTQPTQSRKQLQKREL 120
QY 135 SKASTPKPPCTGVPFAEKVSDPEPQLPEPQPRGRLOHLQASLSQRLGSLDPGWLQRC 194
DB 121 STVPARPPGSKTESCPDEADALPRVPEPRRLQLOQLSSSLRRLTSLDPGWLERC 180
QY 195 HSEVPDFLAPKACRDLGSESSQLIPESAVLGAGS----GQPEASAPQEVSIKRG 250
DB 181 HNRVSDLEVPAGCGLDLSABESQPMGSKVNIADPDIOSEVSVQSPALIAQPAQVLSQ 240

QY 251 SEQPSGGGKRRMNEBEPWESPAVOQESSQAPEPSGAGAVAEDEPGEFVQAQPOP 310
DB 241 SPKSIKNGKRRKNEK-GEPPAQ-DQPSGAGPLSEBARATVHQDPPGEFTQVNVDP 298
QY 311 CSSPSPNPRYHGLSPSSQARAGAEETAPLHFPRLARDKRGYVYVNLNKKYVGRALR 370
DB 299 CNS-----SNQARTERAKGTTHLHSPRASLDRGYVIRLNKKNKRFVGVANR 347
QY 371 SRLRKQAKQKRRKKGCFCGGGATVTTKESCFINEQFDHMAACPRPA-----SEED 424
DB 348 GRLLRKQVKKKKKKKQAAFGSGSPRADTKDTCFRCGQFGHNASQSGPGLTVQEGSD 407
QY 425 TDVAPR-----EPLVSPQVPEVPSLDPTVPLPLYSIGPSGQ 461
DB 408 RDKQPISTLEBVAQRTGASCHSGSEBTQPAAPELQ-VPHCPF---PMSPLYPGPGLGQ 463
QY 462 LAETPAEYFQALFOUGHAFRRGQRAAMRTISGISTLLVPTGAGKSLCYQLPALVYR 521
DB 464 VAETPAEYFQALERGYRFRGQRAIMRLISGISTLLVPTGAGKSLCYQLPALVYR 523
QY 522 RSPCLTLVSPPLSLMDQVSGLPCLKAACIHSGMTKQRESYLOKIRAAQVHLYMTLP 581
DB 524 RSPCLTLVSPPLSLMDQVSDLPCLKAACIHSGMTKQRESYLOKIRAAQVHLYISP 583
QY 582 BALVGA-----GGLPRAQLPEVAPACIDEAHCLSQSHNFRPCYLKVCYLREMGVHC 636
DB 584 BALVCGARGPSLPQAQPLPIAPACIDEVYCHLSQSHNFRPCYLKVCYLREMGVHC 643
QY 637 FLGLTATRTKRAISVVAQHLVAEERDLHGRAPVYTNHLSSMRDQDLTLLOGR 696
DB 644 FLGLTATRTKRAISVVAQHLVAGFEELSGSNITPANHLSVMSRBDQDLVTLLOGR 703
QY 697 FQNLDSIIIVCNRRBETRIALALRTCLAAWVPSGGRAPRTTAAVYAGCSRRRRV 756
DB 704 RRTLDVSIITCTREKIIQNGIMTALRTCLSMGDSRPROCGPALIEAHAGSSGERRV 763
QY 757 QRAFWQGLRVVAVATVAFQMGIDRPDVAVLHLGLPSPSESYVQAVGRDQPARCHL 816
DB 764 QQAFMRGHLRMVAVATVAFQMGIDRPDVAVLHLGLPSPSESYVQAVGRDQPARCHL 823
QY 817 PLQP-----GGEIDR 826
DB 824 FWHQVGSPISPDQRPBGSTIPRLQQLSLCLPVSRCRPGKCGSSVHMTVPMQGBDLM 883
QY 827 ELRRVHADSTDFLVKRLVQRFPACTCTTRPSEBOGAVGGRPVYKYPQOAEOLR 886
DB 884 ELRRVHADSTDFLVKRLVQRFPACTCTTRPSEBOGAVGGRPVYKYPQOAEOLR 929
QY 887 -HQAAPGRRVCMGERALPIQLTVQALDMEBEALITLLCYL-ELPHNMLELATTYTH 944
DB 930 GQQTVPVLGQACLGHERALPVQSTVQALDMEBEATETLLCYLKLPHHMETELLPWYAO 989
QY 945 CRLNCRGPRPAQIALAHNCPRLAVCLAQQLPEDPQSGSSVYFDMVKLVDSMGELASR 1004
DB 990 CHNLCLGSAQIQALAHNCPRLAVCLAQQLPEDPQSGSSVYFDMVKLVDSMGELASR 1049
QY 1005 RALCOLMDHNEBRTGVRRGTVLVEFSELAFLHRSGLTIAEKKQICDPLUGRQVABER 1064
DB 1050 QALHQLKMDPEPKKAAGCTGVLVYFSELAFLHRSGLTIAEKKQICDPLVLRQAHNH 1109
QY 1065 QALARIKRTFOAFHSVAFPSGCPCLQEODEERSTRLKDL 1104
DB 1110 KALAHLQWSKAFRSVAFPSGCPCLQEQNEEHSNQKLV 1149
RESULT 7
QY 09VSE6 PRELIMINARY; PRT; 1579 AA.
AC 09VSE6
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE CG7487-PA.


```

QY 268 -----PWESPAYOQES--SOQAP-----284
      | | | | | | | | | | | | | | | | | |
Db 429 TASGVAVAPVPEINNKVPEPETETFAQENPDPSADEDDQATVPENKKDKAKRQQAOK 488
QY 285 -----PSEGAGAVAVEEDPPGEVQAOPOP---310
      | | | | | | | | | | | | | | | | | |
Db 489 QKTKPKAEPKPKTEKKAKVKAKEKPKAEEKPKRNSKKAIAVEPADPPEDEKQPLNPEDL 548
QY 311 -----CSSPSNPRVH-----GLSPSQARAGKAGTAP 338
      | | | | | | | | | | | | | | | | | |
Db 549 KYVALLEAGDITSVPRIINVOGLEADATQRYITFPAGPNFQSGSITRIVERKRAAR 608
QY 339 IHIIFRLAHD-RGNVYRLNMQOKHYVR-RALRSRLTJKQAMKOKR-----KKG 387
      | | | | | | | | | | | | | | | | | |
Db 609 KLEBRIAAGKLNENPVITINIQKKFKVRGKSVNSKYKQOMRHKRYAVALSGPDMWG 668
QY 388 EFGGGGATVTTKESCFLEQFDPHMAOC-----PRP-----419
      | | | | | | | | | | | | | | | | | |
Db 669 GGDGG---VLT---CFQCGVGHFAQCKVKGDSLPLPSAQLBEDPSPFTLAEAQEMA 721
QY 420 -----ASEEDTDAVGEPLVPSQP-----439
      | | | | | | | | | | | | | | | | | |
Db 722 SOGAVVANSRNTSRLEQANAAILQGDLEINSEBEDQSSGDEVOQHDPNWSDEMDV 781
QY 440 -----VPE---VPSLDP-----VL 451
      | | | | | | | | | | | | | | | | | |
Db 782 DFEALDAVAEASLSQVSOEKASPIKTYVGHKIPBEFLKQAGLDTTASSNRQHGCVK 841
QY 452 PLYSGSPQGLAETPAEVFOALEQUGHQAFRPGQERAVRILISGISTLVLTGAGKSLC 511
      | | | | | | | | | | | | | | | | | |
Db 842 PLYDLFPDSVODTTPVEVLEALHMFHTNFRKGOPRAIKRTLSGLSTLTSGSGSKSLC 901
QY 512 YOLPILVRSRBCPLTLVYSPILSMDQVSGLPCLTKACISGWTQRORSVLOKIRA 571
      | | | | | | | | | | | | | | | | | |
Db 902 YOLPAVLRSKGAITLVISPLSLMEDVGTGFHFLRHCLHTNOGAPORMKILOQIMAN 961
QY 572 AOVHVLMLTPEALV-----GAGGLPFAAQLPVAACIDEAHCLSQMSHNERPCYLRYC 625
      | | | | | | | | | | | | | | | | | |
Db 962 GSIDILVSPBAVAVGERATFGAI--LQQLPPIAFACIDEAHCVSQSHNRPSYLMIC 1019
QY 626 KYLERBMVHCFGLTATRTTRTASDVAOHLAAE-EPDLHGPAVPTNLHLSVMDRT 684
      | | | | | | | | | | | | | | | | | |
Db 1020 KYLRKMLGRTVIGLATAITLPTPRVSIINHLGISDEBERGIIIDIPDLMLVSVSKDENR 1079
QY 685 DQALLTLIOGKRFQNDLSIIYCNREDTERIALRLTGLHAAMPVSGSGRAPKT---T 740
      | | | | | | | | | | | | | | | | | |
Db 1080 DQALLTLIOGKRFQNDLSIIYCNREDTERIALRLTGLHAAMPVSGSGRAPKT---T 740
QY 741 ABAVHAGMCSRRRRVORAFMOGOLRVVAVATVAFGMLDRPVRATLHGLPPSESYVO 800
      | | | | | | | | | | | | | | | | | |
Db 1140 ABAVHAGMCSRRRRVORAFMOGOLRVVAVATVAFGMLDRPVRATLHGLPPSESYVO 800
QY 801 AVGRAGDQOPAHCHFLQPOGEDLRELRHVAHDSITPLAVKRLVQVFPACTCTCTRP 860
      | | | | | | | | | | | | | | | | | |
Db 1200 EIGRGRDQLPCHCHFLDPAKGGDOSELRHVAHDSITPLAVKRLVQVFPACTCTCTRP 860
QY 861 PSEOGAVGERPVPRPQGEABQLSHQAPGR-RVCMGHERALPIOLTVQALDMPERA 919
      | | | | | | | | | | | | | | | | | |
Db 1255 --DKESKRTALPIP-----LEHGDGRVHMCPEHETGSEVKEVTEALDIPAN 1300
QY 920 IETLLCYLELPHHMLLELATTYTHCRILNCPGPAQLOALAHRCPLAYACLQOLPEDEP- 978
      | | | | | | | | | | | | | | | | | |
Db 1301 IETLLCYLELPHHMLLELATTYTHCRILNCPGPAQLOALAHRCPLAYACLQOLPEDEP- 978
QY 979 GGGSSVFEEDMVKLVDSMGWELASVRALCOLQMDHEPTGVARGTGVILVESELAFLIR 1038
      | | | | | | | | | | | | | | | | | |
Db 1361 KEDSNITIEFSVIDIAIGIMNSGVVYQOLKDLLEM--VKVNGYPKSPITVSEYDLGFRIK 1418
QY 1039 SFGDLTAEKDOI CDPLVGRVQAREQALARTQAFHSVAFPSGCGCLEQO--DEERS 1097
      | | | | | | | | | | | | | | | | | |
Db 1419 VEGDFEESI DNALDTLVYRSVKQERTQILOLYAHGLAAVAVSSCGCCNADEPQDRG 1478
QY 1098 TRLKDLIGRYFEEBEGO-----EPGMEDAQPEPQGARLQDWEQVQCDIRQFLSLRP 1151

```

```

Db 1479 EOLKAIIVANYPANDVPQOLELEIEBPSNVPD-----ENIIDVHALINMYP 1523
QY 1152 BEKSSRAVARITFHIGSPCYPAQYVGGDRFWRKYILHSFHALVGLATEBLQ 1205
      | | | | | | | | | | | | | | | | | |
Db 1524 DNTFTGRNIARI FGIIMSPNYPAYWGR-CXKPRAHVAVYDFRILHLNMAIITK 1576

RESULT 8
Q7PNM8 PRELIMINARY; PRT; 1552 AA.
ID Q7PNM8
AC Q7PNM8;
DT 01-MAR-2004 (Tremblrel. 26, Created)
DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE ENSANGP0000004546 (Fragment).
GN Name=ENSANG00000003572;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
RN [1] NCBI_TaxId=180454;
RP SEQUENCE FROM N.A.
RA STRAIN=PEST;
RC Anopheles Genome Sequencing Consortium;
RL Submitted (APR-2003) to the EMBL/Genbank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
   EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is
   preliminary data.
DR EMBL; AAB01008960; EAA11913.2; -.
DR HSP; P15043; 10YV.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006259; P:DNA metabolism; IEA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004589; RecQ.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00098; zf-CCHC; 1.
DR PRINTS; PRO0939; C2HCZNFINGER.
DR TIGRFAMs; TIGR00614; recQ_fam; 1.
DR PROSITE; PS50158; ZF_CCHC; 1.
FT NON_TER
SQ SEQUENCE 1552 AA; 17131 MW; 35DDA746F98A1089 CRC64;

Query Match      26.7%; Score 1713.5; DB 2; Length 1552;
Best Local Similarity 30.0%; Pred. No. 2,4e-73;
Matches 485; Conservative 194; Mismatches 436; Indels 503; Gaps 49;

QY 10 RIQWERAFRRQGRSPQDDVEAAPDEETRALVREYRLTKR-----50
      | | | | | | | | | | | | | | | | | |
Db 15 RVQWKEKFRKKYGRIPSKLDIRESPSEIRMSYKLYOLKTNLHKLTLLEALDDDDPDG 74
QY 75 SLISQASPSQDVSDVSGLLSGTILNDSALLERSSWSGSLTKPIATYQPSDTHFA 134
      | | | | | | | | | | | | | | | | | |
Db 77 -----CNGPHNRAATKSPQPT-----PGRSHQSVVPYQGRILKANLKG--TLQAGP 121
      | | | | | | | | | | | | | | | | | |
Db 135 EKLQDHNLVANQGESKRAVQSTKAEP1EPPK-----PVTINGITLLPST 181
      | | | | | | | | | | | | | | | | | |
QY 122 ALGRPPWP-----LGRASSKASTPKPPTGP-----VPSFAEKYS 156
      | | | | | | | | | | | | | | | | | |
Db 182 ALIPKRNPKRSISRNSPSSQSSST---NGTAPYGVLPDLETLTLTAKSEIESNATKVE 238
QY 157 DEPPQLEPQRPFGQLQHLQASLSQRLSLDPGMQRC-----HSEVDFLAGPACRCP- 210
      | | | | | | | | | | | | | | | | | |
Db 239 EKLPAWP-----LVTPQSEPINQLEDERWINRLSGGTGSDNSVTTTGAPSSSERPT 287

```



```

QY 211 -DLGSEESOLLIPESAVLGPAGSQ-----GPEASAFQEVISIRVG----- 250
Db 288 PTNPELRANVNASLASATSPGMSALSMTDKAGTAVNSQATA---AARIGCVQATAYDA 344
QY 251 -----SPOPSSSG-----GKKRMNEEPWESPANVOQESSQAGPSEBAG 290
Db 345 AIHSDSDGVENSEDESATKTSFGMLHIAKKRKIFTSINRSPE-RVEAKQKPSIETNN 403
QY 221 AVAVE-----EDPG-----EPVQAPR-----QC-----SSPN 316
Db 404 VVSVESSKVENODPCGVABEABEQATPKAPKRVOKRVSKAKOPTKVKRSTTRKSTTIS 463
QY 317 PRYHGLSPSSQARA-----GKAEGTAP-----LHIF 342
Db 464 PK---LSFGSSKPAQGRKAKKIKTKRKVDVGESEBAAPBAEQEYVEPGLARLSI 520
QY 343 PRLAHD-----RG 351
Db 521 PRLDINEIVROTKLADSFILGAGSABEGSSGKTRTTVAGSTKETALRKMAAGKLINE 580
QY 352 NYVRLAMKQKHVVRG-RALRSRLKQAMKQKRRKKEFGG-----GGATVTTKES 402
Db 581 NFVRIDIRKRVFKKKTINYSRYKKSQWK---AKKAALTGPENDMRCDDGTLT--- 633
QY 403 CFLNQDFHMAQCPRPASEE---DTDAVPE----- 431
Db 634 CFQCGGTGHMAQCKKPLEDKLLPYDADTIESSPTLEBAMANSRTLAVHSRIERL 693
QY 432 PLVPSR-----QPVPEVPSLDP----- 448
Db 694 PVNAVPMWKEQEQEDLPAEEVGEVGPKEPSTLPVEDDLDEPMQESDEHOKETEQNAAA 753
QY 449 -----TVLPYISGPGSQALETAEVFOALE 474
Db 754 SOPAPAYIGHKIPBEFLKQSGLEVTWSGGFKGYEPLOYQRCQCTPPEVFEALR 813
QY 475 QLGHOAFRGOBRAMRILSGISTLLVLTPTGAGKSLCYQPLALYLSRSRSPCLTVLVSPLL 534
Db 814 MFGHGSFHHGDBRAMRVLCGLSTLVLTSTGKSLCYQPLALYLRKRSCTTLTVISPLV 873
QY 535 SLMDQVSGLPCLKAACIHSQMTKQKORESVLQKIRAAOVHVMJLTPREALV-----GAG 588
Db 874 SLMEQVHGMPLFNAHCHTNTQTKVNRRTMQALAGVDVLLISPEAVVSGEKSTGCG 933
QY 589 GLPRAAQLPVAFACTIDEAHCLISQWNSHFRPCYLAVCYLRMGVHCHGLTATAT--R 646
Db 934 SL--LROLPPIAFACIDEAHCVSOWNSHFRPSYLMICKLKEKLGKTLGLTATATVQT 991
QY 647 RTASDVACHLAAVEBPD---LHAPRAPVTNLHLSVMDRDTQALLTLQGRFQNDLS 702
Db 992 ROAGSIVSHLAI---PDGVQGIISDIPLDNLLTLTVSRDANDVALVELQSERSSLOS 1048
QY 703 IIVYCNREDETERIALLTCTCLHAAMVPGSGRAPKT---TAEAYHAGMCSRRRRVQ 757
Db 1049 IIVYCTRDDCERVAFTIRTCQDABAAAANAANKRKLAVABEYHAGMASRRRTIQ 1108
QY 758 RAFWOGLRVVAVTAFAFGMGLRDPVRAVLAHLGLPSPFSYQAVAGRAGDCQAPACHLF 817
Db 1109 NAFMGEGLRIYVATTIAFGWINKADIRAIHYNMPNFESYQVEVGRAGRDILSHCHF 1168
QY 818 LQPOBEDLRELBRRHNDSTDLAVKRLVQVRPACTCTPSPSQ-----EGAVGGR 872
Db 1169 LDGKGTDRNELRRFTYANSIDHVIKRLQKIFVPACQ---KIQOQVILLFTDGGGGGGR 1224
QY 873 PVPKYRQEBASQLSHQAAGPRVVCGERALPIQTLVQALDMPBEAIFTLLEYLH 932
Db 1225 -----KRLCGHICFSIATVQOQLODIPRENNITPLCYLELDQ 1263
QY 933 HWLELLATTYTHCRINCPCGPAQOLALAHRCPLAVCLAQOLPEDPGQ-SSSVEFDWYK 991
Db 1264 RYIQALSPAYTWCKWMSYGVVRPLROAKACECPPLAMAFALDLKRGISHASTAIEPVID 1323
QY 992 LVDSMGWELASVRALCOLQOMDHERPTVGRGTGVLVERSBELAFHRSGLDLAEKQDI 1051

```

```

Db 1324 VASATGMDSGVVKQKOLKEW---TTVNVNRKRSPLSVFTELGFRVRAPCDLTEBELDHA 1381
QY 1052 CDFLYGRVQAEERQALALRRTFOAFHSAVAPSCPCLEQOD--BEETRKLXDLGRYFE 1109
Db 1382 LDGLYERTYHBERQALQLOLYISDLNSVCENTIGP-VSRADCPGPSDKKTYIREYF 1440
QY 1110 EBEQEPGCMEDAOGEPEQARLDQW-----EDQVRCDIROPLS 1148
Db 1441 TDISKE-----QIEIVPEVISIMIKRFXELAVTVIPLQPDTTTBEQLINDIRTAIC 1493
QY 1149 LRPEKFSRAVARIFHGIGSPCYPAQYVGGDRFRWRYLHLSFVALGLATELLQV 1206
Db 1494 RYFENNFTGRALARLFHGVQSPNTALVMSRS-NFWRAYTQTDENRIVRLANABIVRM 1550

RESULT 9
Q9NH11
ID Q9NH11 PRELIMINARY; PRT; 1530 AA.
AC Q9NH11
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE RECQ4.
OS Name=RecQ4; Synonyms=RECQ4;
OC Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Kusano K., Barres M.E., Engels M.R.;
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF233659; AAF2939.1; -.
DR HSSP; P15043; IOY.
DR FlyBase; FBgn0040290; RecQ4.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0016787; F:Hydrolase activity; IEA.
DR GO; GO:003676; F:nucleic acid binding; IEA.
DR GO; GO:006259; P:DNA metabolism; IEA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004589; RecQ.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00098; Zf_CCHC_1.
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC_C; 1.
DR SMART; SM00343; ZNF_C2HC; 1.
DR TIGRFAMs; TIGR00614; recQ_fam; 1.
DR PROSITE; PS50158; ZF_CCHC; 1.
KW ATP-binding; Helicase; Hydrolase.
SQ SEQUENCE 1530 AA; 169876 MW; 1B753A59A3B30EB4 CRC64;

Query Match 32.2%; Score 1680.5; DB 2; Length 1530;
Best Local Similarity 26.4%; Pred. No. 8-8e-72;
Matches 420; Conservative 191; Mismatches 412; Indels 275; Gaps 36;

QY 107 QRLKANLKGTLQAGP-ALGRBPWPLGRASSKASTPKPGTGVPSFAKVSDEPQLBPP 165
Db 306 QELQTDNMSNMWQKQDHLNHTP-PASQKSVAPKKKP-----PEQGTSDSDSDEVAES 359
QY 166 QPRBRILHLQASLSQRLGSLDPCWLO-Q-RCHSEVVDPLGAKKAC-----RPLGSEESQ-- 218
Db 360 BEEQEPQRYOLSKRKRTVSTASGVAAVAPVEIPNKVPEPTETFTFAQENPQFSADEDDA 419
QY 219 LLIP-----GESAVLGPAGSQ-----GPEASAFQEVISIRVGSQPSGSGEKKRMNEEPWES 271

```

```
Db 420 TVPENKKKAKRKOAGKOKTKPKA-----EPRKKT-----EK 455
Qy 272 PAVOOESSQAG--PPSEGAQAVAEDEPPGEVQAOQPP-----CSSPSNR 318
Db 456 KAKVAKAEKKPKAKBKPRNSKKAIIVAPADPEDEKQPLNBDLKVTLAEAGDITSVR 515
Qy 319 YH-----GLSPSSQARAKAGTAPLHIPPRLARHD-RGNY 353
Db 516 INVQDLBEADATAGRYITRTFAAGPNPGLSEGSNIRIDEBRAAKKLEERIAAGKLENEF 575
Qy 354 VRLANKOKHYVNG-RALRSRLLRKQAWKOKMR-----KKGCGGCGGATVTTKESC 403
Db 576 VVINIOKKKFFVKGKSVNSFKYKQOQWHRKKRVAAALSGPDMMDGGCGG---VLT---C 628
Qy 404 FLNEQPDHWAOC-----PRP----- 419
Db 629 FCCGGVGHFAOQCKYKKGDSLPLSLAOLBEDPBPPTLAEAOEMAGAVVAHSRINSLRP 688
Qy 420 -----ASEBDTAVGPEPLVPSPQP----- 439
Db 689 QANANAILQGDLENESEEDQJESSDEEVQGHDPNWSDEMDVDPEALDAAVEASLSQP 748
Qy 440 -----VPE---VPSLDPT-----VLPYLSGPGGQLAETPA 467
Db 749 VSOEKAASPIKTYVGHKIPDEEPLKQAGLDTTASSNSRQHGKPKLYDLLPDGGSYQDTTP 808
Qy 468 EYFOALEQGHQAFRRGGERAVWRLISGSLTLVLP.TGAGKSLCYOLPPLALYSRRSPCLT 527
Db 809 EYLEALHMFHGTNFRKGQORALTRTSLGSLVLTSLTSGSKSLCYOLPPLALYSRKQALIT 868
Qy 528 LVVSPPLSLMDVQSGLPCLKAACISGKMTKQKQRESVQKIRAOVHVLMTPEALV-- 585
Db 869 LVISPLVSLMEQVGVPHFLRAHCHTNOTAPQBRKIQOMLANGSIDLILVSPBAVNG 928
Qy 586 ----GAGLPRAPOALPPVAPACTIDEAHCCSOWSHNRPCTLYRCKTLRBMGYHCLGIT 641
Db 929 ERATGFGAI--LRQLPPIAFACIDEAHCVSOWSHNRPCTLYRCKTLRBMGYHCLGIT 986
Qy 642 ATATRTASDVAGHLAVAE-EPDLHGPARVPNTLHLSVMDRDTQALLTLLOGRPOUL 700
Db 987 ATATLTPTRVSLTNHGLISDGERGILSDPLPDNLVLSKDENRDALLQLNSEFEPC 1046
Qy 701 DSIILYCNREDTERIALLRCTCLHAAMPSSGSGARPKT-----TAEVAHAGMSRRRRV 756
Db 1047 QSIILYCTRDEBERIAGRICTVQDRREPTQDTKKRQVWQKQEPHAGMPASRRRY 1106
Qy 757 QRAFMQGLRVVVAIVAFGMDLRPDVRAVLHLGPPSESVYQAVGARGDQPAHCHL 816
Db 1107 QRAFMNSNEIRIVAVATIAFGMGINKPDIRAVIHYNMRFNPESVYQOIGRAGRDGLPSCHL 1166
Qy 817 FLPOQGEDRLRLRHVHADSTDFLAVKRLVQRFPACTCTCTRPSEOGAGVGERPVK 876
Db 1167 FLDAKGGDSELRHHYNSIDHVRKLLQKIFVCSG-----DKRASKRTALPIP- 1218
Qy 877 YPPOAEQLSHOAPRPR-RVCMGHERALPIQLTVOALMPEEALITLLCYLELHPHML 935
Db 1219 -----LEBGSFRVHMCBGHIGSVKTYEMDIPRAENITLTLCTYELDRKCI 1267
Qy 936 ELLATTYTCRLNCPGSPAOLOALAHRCPLPCLVLAQOOLPEDP-QGSSSSVEFDWKLVD 994
Db 1268 SVLSAYWMAKVSYSGPKYLKHAACECPPLAMAILQIRDKTFKEDSNLIEFSVDIA 1327
Qy 995 SNGWELASVRALCOLOMDHERPTGYRKGTVLVESELAFLHRSQDLTAEKQOICP 1054
Db 1328 GIGWNSGVVXYQKDLDM--VKVNGYPKSPITVSVYDLGFRILKVGDPDTESIDNALDT 1385
Qy 1055 LVGRVQARERQALARTTQAFHSAVAFSCGPCLEQO--DEERSTLXOLLGYYFESEEG 1113
Db 1386 LITRSVKQERTQILQIQYAHGLAAVAYS CGCCCAADPADPDGBQDLKAIVRNYFANDYP 1445
Qy 1114 Q-----EPGMEADAGPBPGRQARLQDWEQVCDIRQFLSLRPEKFSRAVARLPHGI 1167
Db 1446 QDLELEIEBSNVDP-----ENIIDVHALIMNYPDNTFTGRNIRARLPHGI 1490
```

```
Qy 1168 GSPCYPAQVQDRRFRKRYLHLSFHALVGLATEBLQ 1205
Db 1491 MSPNYPVITWGR-CKFWRAHYVVDNFRILHLANMAIILK 1527

RESULT 10
ID Q7YR85 PRELIMINARY; PRT; 340 AA.
AC Q7YR85;
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE RecQ protein-like 4 (Fragment).
GN Name=RECQ4;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
OX NCBI_TaxId=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14667821; DOI=10.1016/S0888-7543(03)00238-6;
RA Winter A., Alzinger A., Fries R.;
RT "Assessment of the gene content of the chromosomal regions flanking
RT bovine DGAT1."
RT Genomics 83:172-180(2004).
RT EMBL: AJ518973; CDS58810.1; -.
DR HSSP: PL5043; 10YF.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF00271; Helicase_C; 1.
DR SMART: SM00490; HELIC_C; 1.
FT NON_TER 1
FT NON_TER 340
SQ SEQUENCE 340 AA; 37366 MW; 205F057DD109CAED CRC64;

Query Match 18.2%; Score 1172; DB 2; Length 340;
Best Local Similarity 67.8%; Pred. No. 2.7e-48;
Matches 232; Conservative 27; Mismatches 57; Indels 26; Gaps 2;

Qy 601 FACIDBAHCLSGWSNRPCTLYRVCKV-----LREBMGYHC 636
Db 1 FACIDBAHCLSGWSNRPCTLYRVCKVQVSLGKGSVQGRBGTAGSPPSQTLRQMGVHC 60
Qy 637 FLGLTATATRTASDVAGHLAVAEPPDLHGPARVPTNLHLSVMDRDTQALLTLLOGR 696
Db 61 FLGLTATATRTASDVAGHLAVAEPPDLHGPARVPTNLHLSVMDRDTQALLTLLOGR 120
Qy 697 FQNLDSIIYCNREDTERIALLRCTCLHAAMPSSGSGARPKTTLAEVAHAGMSRRRRV 756
Db 121 FQNLDSIIYCNREDTERIALLRCTCLCDADPPHGLAEVAEVAHAGMSRRRRV 180
Qy 757 QRAFMQGLRVVVAIVAFGMDLRPDVRAVLHLGPPSESVYQAVGARGDQPAHCHL 816
Db 181 QRAFMNSNEIRIVAVATIAFGMGINKPDIRAVIHYNMRFNPESVYQOIGRAGRDGLPSCHL 240
Qy 817 FLPOQGEDRLRLRHVHADSTDFLAVKRLVQRFPACTCTCTRPSEOGAGVGERPVK 876
Db 241 FLPOQGEDRLRLRHVHADSTDFLAVKRLVQRFPACTCTCTRPSEOGAGVGERPVK 298
Qy 877 YPPOAEQLSHOAPRPR-RVCMGHERALPIQLTVOALMPEE 918
Db 299 AVPAASADQSGQSPVHTPRCGHERVLPVQPTVOALMPEE 340

RESULT 11
ID Q7X829 PRELIMINARY; PRT; 927 AA.
AC Q7X829;
```



```

DR HSP: P15043; 10YY.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO: GO:0016787; F:nucleic acid binding; IEA.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR GO: GO:0006259; P:DNA metabolism; IEA.
DR Pfam: PF00271; DEAD; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
DR TIGRfam: TIGR00614; recQ_fam; 1.
DR PROSITE: PS00690; DEAD ATP HELICASE; UNKNOWN_1.
DR ATP-binding; Helicase; Hydrolase.
SEQUENCE 870 AA; 96382 MW; 8BSADA4F7A81723 CRC64;

Query Match 14.0%; Score 897.5; DB 2; Length 870;
Best Local Similarity 26.7%; Pred. No. 8.6e-35;
Matches 295; Conservative 160; Mismatches 368; Indels 281; Gaps 38;

QY 133 ASSKASTPKPGGVPVFAEKVSDERPDLPERPGRLOHQAISLQSLDPCWQ 192
DB 3 SSSSSSKPRAP--THPPNPQEAIVPSPYPPPPSP-----LFTNLPRIQSDPA--- 53
QY 193 RCHSEVPDLGAPKACRPDLGSEESQLIPGESAVLPGAGSGCPASAFQEVSRVQSP 252
DB 54 RFSSSVSF---SRLCRSFSTVEKLKSDGVDFV-----PEPLVEVIA----- 95
QY 253 QSSSGGKKRRNNEEFWESPAPVOQESSQAGPPSBGAVAVEBPPG---EPVQAQPPQ 309
DB 96 -----PPK-----SVRRKPPNITDTITSPYK 118
QY 310 PCSSPBNPRYHGLSPSSQAPAGAEGTAPLHIFPLARHNRGNVYLNKKOHVYGRAL 369
DB 119 PWFPSN-----GNGE-----GNFVNLNNGK---RGKPF 145
QY 370 RSRL-----LRQAWKQKMRKKGECFGGATVTTKESCFINEQFDHWAACPPRASE 423
DB 146 PKYKGVSKRSRYSRGRKRYKKKADGGESELL--EESSDIQKQ-----ED 191
QY 424 DTDANGPEPLVSPQCPVEVPSIDPTVPLVYSLGSPGQLAETPAEVFOALEQHQAPR 483
DB 192 EANGF-----ISSVEDILAVKTEASDENITKILNIVY-----GYDSFRD 231
QY 484 GGBRVAWRLISGISTLLVPTGAGSLCYQALPALYSRSPCLTIIVSPSLIMDDVSG 543
DB 232 GOLQATKMLGSSSTLVPTGAGSLCYQIPAMII---PGITLVSPVLSIMDDQHN 287
QY 544 LPPLCLACIHSQMTKQRESVLQKIRAAQVHMLTPALVAGAGLPAAQLPVAFAC 603
DB 288 LPSIIKGLSSQREBEATELRKKEGIIKVLFPVSPRLNVEFLSMFRMSLSLVV 347
QY 604 IDBAHCLGWSHNPFCYLRV--CKVIREMGVHCFLGTATATRRPASDVAOHLVAEPP 662
DB 348 VDEAHCVSESHNFRSPYMKLAKSMFLSEKKECIAMATATMTQLQVMSLEI--PST 406
QY 663 DLHGPAVPVNTNLISVM---DRDTPQALLTLQGRFQNLISIIYCCRRRETEIAL 719
DB 407 NLIQSQLDNDELSTLSGANMKD--LILIMESPVEIKISIIYCKQFETDMISYK 464
QY 720 LFTCLHAAPVSGSGRAPKTTAEVYAHAGCSRERRVQAFQMGQQLRVVAVAFMGMD 779
DB 465 LND-----NNINAKGYSHGLPAKDRVAIQSFSCNKRIVVAVATVAFMGMD 510
QY 780 RPDVRAVHLGLPSESVYQAVGARGDQPAHCHLFLQPOGEDILRELRHVHADSTP 839
DB 511 KGVGVAVIHFSVPGSEWEEYQVIGRAGDRLSYCHLFY--DNDTYLKLRLSHSDVD 568
QY 840 LAVKRLVGVFRACCTCTRRPSEGGAGGERPVKYPQAEQSLSHQAPRRVCMG 899
DB 569 YAAKGLTIVP-----STEVQHE-----KIC-- 590
QY 900 HRAALPIQLVQALDMPBEAIEITLCYELHPIHMLIETATVTYHGRNLCPGGPQLQAL 959

```

```

DB 591 ---SLVIESAQKEDMKKEEVMQITLHLEGEVQYLRMLPOLNICTLIN----- 636
QY 960 AHRCEPLAVCLAQQLPEBPGSSSVFERDMKLVDSWG---WELASVARALCQLQWDEP 1016
DB 637 FHKSSPNTLAAR-----SATVAAILKSHVQGLHVPDIPAVASSICVAT----- 681
QY 1017 RTGVARGTGVLFVEFSELAFLHRSPPDLTAEEKD-----QICDFLVGRVQABER 1064
DB 682 -----IDVLAIEIOA-----LKMGEVITYELKDSAFCYTILKSPKEICSLSHLKWLTE 730
QY 1065 QALARLR--TFOAFHSVAPPCGCPLEQDEBERSTRKOLLGRFYEEBGEPPGMDAQ 1123
DB 731 IESCVRKLDIMSSAAVAIVSNTSELSGAKQTR--SLOSRIIDYFNG-----DEK 781
QY 1124 GPEPGQALQWEDQ-----VPCDIHQFLSLRPEEKSSRAVARIPIHGISPCYPAQYGO 1179
DB 782 CDSPSKA-----TQNCAPLRADIVPFIQSNRQAKFTPALRIHMGVSPAFNSVMSK 835
QY 1180 DRFPFRKYLHLSFHALVGLATEEL 1203
DB 836 -THFWGRYMNVDFRYIMEAAQTEL 858

```

```

RESULT 13
09CGN0
ID 09CGN0 PRELIMINARY; PRT; 941 AA.
AC 09CGN0;
DT 01-UN-2001 (TrEMBLrel. 17, Created)
DT 01-UN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE ATP-dependent DNA helicase RecQ, putative.
GN Name=F28L5.4;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OC NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,
RA Maiti R., Romling C.M., Koo H., Fujii C.Y., Utecherback T.R.,
RA Barnstead M.E., Bowman C.L., White O., Mierman W.C., Frazer C.M.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC079280; AAG50580.1; -.
DR HSP: P15043; 10YY.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO: GO:0016787; F:hydrolase activity; IEA.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR GO: GO:0006259; P:DNA metabolism; IEA.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR011545; DEAD/DEAH_N.
DR InterPro: IPR002464; DEAH_box.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR004589; RecQ.
DR Pfam: PF00271; Helicase_C; 1.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
DR TIGRfam: TIGR00614; recQ_fam; 1.
DR PROSITE: PS00690; DEAD ATP HELICASE; UNKNOWN_1.
DR ATP-binding; Helicase; Hydrolase.
SEQUENCE 941 AA; 104155 MW; 30BC454C795522E1 CRC64;

```

```

Query Match 13.6%; Score 872; DB 2; Length 941;
Best Local Similarity 26.2%; Pred. No. 1.5e-33;
Matches 305; Conservative 165; Mismatches 390; Indels 306; Gaps 43;

QY 93 TGRGRQGVPRPYGRLKANKGTIQAGPALGRRPWPLGRASSKXSTPKPGGVPSPA 152
DB 15 TPRPD---SPFSSPQLOSRAHV---PPYSRKM-----TSSSSKPRAP--THPPNVS 62
QY 153 EKVSDPEQLPEPQRPRLQHQAISLQSLDPCWQORCHSEVPDLGAPKACRPDL 212

```

```

Db      63 QBAVPSPYPPPPPPSP-----LFTNLPRICOSQPA---RFSSVSSF---SRLCSRAS 111
Qy      213 GSEBSOLLIRGSATVPGAGSGCPGASAFQEVSIIVSGSPQSSGCKRNMEEWESP 272
Db      112 FTSVEKLKSDGVDFV-----PPPLVEVIA----- 136
Qy      273 AVOOESSQACPPSEBAGAVAVEEDPPG---EPVQAPQPPCSSPSPNRYHGLSPSSQAR 329
Db      137 -----PPK-----SVRRKPMLITDITSPYKPMVFRSN----- 166
Qy      330 AGKAEETADLHIFRLAHDRGNVYLNKKQHYVGRALRSFL-----LRKQAWKQK 383
Db      167 -NGGE-----GNFVXLNNGK---RGKFPSPKXKGVSKSRSSSPRGKR 206
Qy      384 RKKGECFGGCAVTTKESCFLNEQFDHMAQCPRASEEDTDVAVPEPVSPQVPEV 443
Db      207 YKKKEDAGGSESL--EESSDLOKQI-----DEANGF-----I 238
Qy      444 PSLDPTVLPLYSIPSGQLAETPAEYFOALEQGHQAPRGQBRAMRIISGISTLVLP 503
Db      239 SSVEDAILAVKTEASDENLTKLNLVY-----GYSPFDGQALIKMLIGSSITVLVP 292
Qy      504 TGAGSLCYQPLALLYSRRSPCLTLVSPPLSIMDQVSGLPCLKAACISGMRKORE 563
Db      293 TGAGSLCYQIPAMIL-----PGITLVSPVLSIMDQKHLPSIIKGLLSSQPREAT 348
Qy      564 SYLQKIRAAQVIVMLTPREALVAGGLPRAAOLPPAFACIDEANLSC----- 612
Db      349 ETLRLKKEGIIKVLFPSPERLNVFSLMFRMSLSVSLVAVDEAHCVSICVLSLFCUS 408
Qy      613 W-----SHNRPFCYLRV--CVLYBERMGVHCFGLTATARTASDVQAOLVAVEED 663
Db      409 WLLRALIRNSHFRPBYMLKASMLFSELKACITLMTATMTITQAVMSLIEI-PSYN 467
Qy      664 LHGPAVPPTNLHLSVSM---DRDTPQALLTLQCKRFQMLDSIIICNRRDEPTERIALL 720
Db      466 LIQKQGLRNFELSVLSGANRMKD--LIILMESPYYKEIRSIIVYCKQOYEIDMISKYL 525
Qy      721 RCLHAAMVPGSGGAPKTTAAEYAHAGMCSREERRVQAPFMQGLRVVAVATVAFGKGLR 780
Db      526 RD-----NNINAGYHSGLPADKQVRVIOESFCSKIRVVAVATVAFGKGLDK 571
Qy      781 PVRAVILHGLPPSPFESYQAVAGRDGAPAHCHFLDPQGEDLELRLRHVADSTPL 840
Db      572 GIVGAVIHSVPSGHEEIVQEIQRARQDRSLCHLFLY--DNDTYIKLMSLAHSDVDY 629
Qy      841 AVKRLVQVFPACTCTCTRPSPSEBAGVGERPVPKYRQBAEQLSHQAPGRVRCMGH 900
Db      630 AVGKFLTHYF-----STETKQHE-----KIC-- 650
Qy      901 EBALPIQLTVQALDMEBAIEFLLCYELHPHHMLLELATYTHCHLNCPPGPAQLQALA 960
Db      651 --SLVIESASQKFDMEEMQOTILTHELGEVOYLMLPOLNICCTLYN-----F 697
Qy      961 HR---CPPLAVCLAQOLPEDPGQSSSVFDMVKLVDSWG---WELASFRALCOLQMDH 1014
Db      698 HTTSDAQNLVMLYVQSSPNTLAARSATVAIILKSHVKQGLHVFDPAAVASSICVAT-- 754
Qy      1015 EPRTVGRGTGVLVSESLAFHLRSPGDLTAEBKD-----QICDELXGVQAR 1062
Db      755 -----TDVLAELQNL-----KGEVTEHLKQSAFCYITLKSPEKICSLSHHTKML 799
Qy      1063 EHQALARLR--TFQAFHSAVAPSCPCLEQOQDEBERSTRKLDLGRYEEBEOQEPGQMD 1121
Db      800 TEIESCKYAKLDIMSSAAVAIVSNTSELSGAKQTR--SLOSRIFFDFNG-----D 850
Qy      1122 AGPPEGCARLDQWEDQO-----TRCDIROFLSLRPEKFSRAVARIFHIGISPCVPAQVY 1177
Db      851 EKCDSPSKA-----TONCAFLRADIKVFLQSNRQAKFTPRAIRIMHGVSPPAFENSVW 904
Qy      1178 GQDRRFWRKYHLHSFHALVGLATEEL 1203
Db      905 SK-THFMGKYNMVDPRVIMEAQTEL 929

```

```

RESULT 14
ID Q84RP1 PRELIMINARY; PRT; 899 AA.
AC Q84RP1;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Similar to ATP-dependent DNA helicase.
GN Name=H0302E05.4;
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhacoidae; Oryzae; Oryza.
OX NCBI_TaxId=4530;
RN (1)
RP SEQUENCE FROM N.A.
RX PubMed=14693376; DOI=10.1016/j.gene.2003.09.008;
RA Wang R., Hong G.F., Han B.;
RT "Transcript abundance of rml1, a putative Cfl-1-like factor in rice, is
RT up-regulated by Magnaporthe grisea and down-regulated by light.";
RT Gene 324:105-115(2004).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=22780373; PubMed=12897975;
RA Wang R., Hong G.F., Han B.;
RT "Characterization of the copy number of RIR10 retrotransposon and
RT analyses of transcriptional activity of its LTR in rice genome.";
RT Acta Biochim. Biophys. Sin. 35:768-773(2003).
DR EMBL; AL627350; CAD79701.1; -.
DR HSSP; P15043; 10YV.
DR Gramene; Q84RP1; -.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0006259; P:DNA metabolism; IEA.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR004589; RecQ.
DR Pfam; PF00270; DEAD; 1.
DR SMART; SM00487; DEXDC; 1.
DR TIGRFAMs; TIGR00614; recQ_fam; 1.
KW Helicase.
SQ SEQUENCE 899 AA; 99340 MW; 48D6A843F055749C CRC64;

Query Match 12.0%; Score 772; DB 2; Length 899;
Best Local Similarity 23.8%; Pred. No. 8.3e-29;
Matches 286; Conservative 165; Mismatches 390; Indels 362; Gaps 41;

Qy      40 ALYREYRLTKRTTGGAGGGLSSSESLPAAABEAPRCMGPHLNAATKSPQTPGRSRQ 99
Db      16 AFFLARLRIRCTPPAHAAADSDSGSLSDVSASPPRRSP--LRPAPPPPPPPKGR- 72
Qy      100 GSVVDYGRKLANLKGTLQAGPALGRPMPGLRASAKASTPKRPCTGVPVSPFAKVSDEP 159
Db      73 -----TRPAAPTKPLKLTTPAASAPAP 97
Qy      160 POLPEPQPRGLQLQASLSQRLGSLDP--GYLQCHSEVPDFLCAPKACRPDLCSBS- 217
Db      98 P--PPPTLR-----AAALS-----DPHGIAAR-----IAAGSALTAASGASS 134
Qy      218 --QLIIPESAVLAGGAGSQGPEASAFQEVSIIVSGSPQSSGCKRNMEEWESPQAV 275
Db      135 SFRRLVQSRNBSFDATFTAPASAPSEVSAAPRPPTAA-----TTAPROT 183
Qy      276 QOESSQACPPSEBAGAVAVEEDPPGPVQAOPOPCCSSPNRXYHGLSPSSQARAKAEG 335
Db      184 RPKRVHPSVSEVAASAABQP-----KRAAG----- 212
Qy      336 TAPLHIFRLAHDRGNVYLNKKQHYVGRALRSRLRKQAWKQKRGKGCFCGGGA 395
Db      213 -----SEGNFVRINI--NGYGRRTFKNSQAKSTKCRSWRQ---RAAGA 253
Qy      396 TVTTKESCFLNEQFDHMAQCPRASEEDTDVAVGEPLVPSPQVPEVPSLDPTVLPLYS 455

```

Db	254	T-----	PROGDSBGDLVAAELERKQASD-----	SVLEAVE	287
Qy	456	LPGSGQLAETPAEYFOALBOL-----	GHOAFPPGGERAVMRLISGISTLLVPTGAGSL	510	
Db	288	-----SVAREPSE--QNLKSLINAAVGHDSFPGQOLEAIQQIVAGESTVLVPTGAGSL		340	
Qy	511	CYOLPALLYSRRSPCLTIYVSPFLSLMDQVSGLPPLCKAAICIHSGMTKQKQESVLOKTR		570	
Db	341	CYQVPAFMLT---PBLTIVSPFLSLMDQLEKLPFLFGGLASSQTSDEPHDTLORLR		396	
Qy	571	AAQVHVLMTREATAGAGGLPPMAOLPPVAFACIDEAHCLASOMSHNFRPCYLRV--CKVLK		629	
Db	397	AGEIKVLVSPERFLNEEBFLIFRDTLPISVAIDEAHICISMHNFRSYLRRLASLLR		456	
Qy	630	ERMGVHCFGLTATATRTRTASDVACHLAAVEEPDLHGPAVPFTNLHLSVMDRTDQALL		689	
Db	457	RKLNVQCIAMATATATQTOLEIEIMNALLEIPSD-NLIQTSQIRENLIQTSISTSDNRKIDML		515	
Qy	690	TLLOGKRQONDSIIYCNRRREDTERIALLTCTCLHAANVPSSGGAAPRTTAAEAVYAGNC		749	
Db	516	LLLSPPFVDMKSIIVYCKFOAETPVSXYL---CQD---NITASVHSGLL		561	
Qy	750	SRERRVORAFMGOGLRVVATVAAGMGLDRPDAVLHGLPPSESYQVAVGRAGDG		809	
Db	562	IKNSRRVQELPSCNKKIRVF-----	SL-----	582	
Qy	810	QPAHCHLFLQPGEDLRELRRHVADSTFDLAVKRLVQVFPACTCTCTRPSEBOGAVG		869	
Db	583	-----NFTQEP-----	FLGSDGVGVAMSKFLYQIF-----	SSENTTG	615
Qy	870	GERPVPKTPQGEALSHOAGPPRRVCMGHERALPIQITVQALDMPERAEITLCTYLL		929	
Db	616	-----	CIG-----SLAKELTSRKFDIKEVLLTILYOLEI	645	
Qy	930	HPHHLELATYTYTCRLNCPGPAQOLALAHRCPIAVC-----	LAQQLPEDPGQSSSV	985	
Db	646	GDQYIRLLPQSVYCTL-----	YFKITSQQLADNDILRSVLANSEMDGHY	694	
Qy	986	EFDVAVKLVDSMGWELASVRRALCOLMDHEPRTGVRGTGVLEFS--ELAFHLRSPG---	104		
Db	695	VFDIPRIANDLEKITMNEV-----	FDHLHK-----	LKFGSGISFELKDPAYCY	736
Qy	1042	-----DLTAERKQOICDFLYGRQVARRQALRLRTFOAFHSAVAFSC-----		108	
Db	737	VILMRPDPFNALSN-----	LTKWLSSEVSSKISKLDAWF--ALANAIVGCKKRTGCGSGS	790	
Qy	1086	--GPCLEOODERSTRKLDGLRYFEEBEGQEPGGMEDAQGPBGARQIOWMEDOYRCDI		114	
Db	791	QHTPIQCKIME-----	YRSKDGTSBNC-----	RTQLOKSSPFLQADI	830
Qy	1144	ROFLSLRBEKFSRAVARIFIGISPCYPAQVYGGDRRPMRKYLHLSPHALVGLATEEI		120	
Db	831	KVFIQSNFPAKFTPRAVARIMHGISPAFPSTWSKN--HFWGRYVEVDPLVVEAKAEL		889	
Qy	1204	LQV 1206			
Db	890	VKL 892			
RESULT 15					
Q7FAHO PRELIMINARY; PRT; 874 AA.					
AC Q7FAHO;					
DT 05-JUL-2004 (TREMBLrel. 27, Created)					
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)					
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)					
DE OJ000223_09.17 protein.					
GN Name=OJ000223_09.17;					
OS Oryza sativa (Japonica cultivar-group).					
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;					
OC Ehrharioideae; Oryzeae; Oryza.					
NCBI_TaxId=399477;					

[illegible]

```
Qy 757 QRAFMGOLRVVAVATVAFMGGLDRPDRAVLHLGLPSPESYVQAVGRAGRDQPAHCHL 816
Db 544 QELFCSNKIRVF-----SY-----NF 559
Qy 817 FLQPGEDLRELRRVHADSTDFLAVKRLVQVFPFA-----CTCTRPPSBOGAVGE 871
Db 560 TEOP-----FIGSDGVGDYAMSKFLYQIFSESENTTGCISLAK-----597
Qy 872 RVPVKYPPQEAQLSHQAAFGPRRVCMGHERALPIQLTYQALDMPBEALETLLCYIELHP 931
Db 598 -----ELTSRKPDIKEVLLTTLTQLEIGD 622
Qy 932 HHMLLATTYTHCRINCPGPAQLOALAHRCPLAVC---LAQOLPEDPGSSSVVF 987
Db 623 QOYIRLLPQFSVYCTL-----YHKTSPLQILDADKDLIRSVLNRSEMKDGHYF 671
Qy 988 DMVKLVDSMGWELASVRRALCOLMDHEPRGTGVRGTGVLVEFS-ELAFHLRSPG-----1041
Db 672 DIPRIANDLKITMNEY-----FDHLHK-----LKFSGEISFELKDPAYCVI 713
Qy 1042 -----DLTAEKDQICDPLYGKVGQARERQALRLRTFOAFHSVAFPSG-----1085
Db 714 LWRPDDFNALSAN-----LTKWLSVESSKISKLDAMF-ALANFAVKGCKRTGCGSGSQH 767
Qy 1086 GPCLBQODERSTRKLDLGRYFEBEEOGEPGMEDAQGPBPQARLODMEDQVRCDIRQ 1145
Db 768 TPCIOKKIME-----YFSKDDGTSENDG-----RTQLOKSSPFLQNDIKV 807
Qy 1146 FLSLRPEEKFSRAVARIFPHIGSPCYPAQVYGODRRFRWKYLLHLSFHALVGIATEELLQ 1205
Db 808 FIOSNSPAKFTPRAVARIHIGISSPAFSPVTSKN-HFWGRYVEVDPLVMEAKAKELVK 866
Qy 1206 V 1206
Db 867 L 867
```

Search completed: March 19, 2005, 01:08:06
Job time : 210 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 19, 2005, 00:58:06 ; Search time 46 Seconds
(without alignments)
1960.349 Million cell updates/sec

Title: US-09-889-325-4

Perfect score: 6424

Sequence: 1 MERLDVRRRLQAWERAFRR.....HLSFHALVGLATEELLQVAR 1208

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA: *
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep: *
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep: *
3: /cgn2_6/prodata/1/1aa/6A_COMB.pep: *
4: /cgn2_6/prodata/1/1aa/6B_COMB.pep: *
5: /cgn2_6/prodata/1/1aa/6C_COMB.pep: *
6: /cgn2_6/prodata/1/1aa/Backfill1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6424	100.0	1208	US-09-463-702A-2	Sequence 2, Appl1
2	6424	100.0	1208	US-09-699-135-2	Sequence 2, Appl1
3	1883	29.3	361	US-09-463-702A-37	Sequence 37, Appl1
4	1883	29.3	361	US-09-699-135-37	Sequence 37, Appl1
5	632.5	9.8	607	US-08-781-891-75	Sequence 75, Appl1
6	632.5	9.8	607	US-09-618-166-75	Sequence 75, Appl1
7	621	9.7	370	US-08-559-303B-77	Sequence 77, Appl1
8	621	9.7	370	US-09-175-828-77	Sequence 77, Appl1
9	621	9.7	370	US-09-753-143-77	Sequence 77, Appl1
10	613.5	9.6	334	US-09-463-702A-38	Sequence 38, Appl1
11	613.5	9.6	334	US-09-699-135-38	Sequence 38, Appl1
12	606	9.4	619	US-09-489-039A-8075	Sequence 8075, Ap
13	591.5	9.2	641	US-09-543-681A-4359	Sequence 4359, Ap
14	585	9.1	1394	US-09-502-540-16497	Sequence 16497, A
15	554.5	8.6	589	US-09-134-000C-5710	Sequence 5710, Ap
16	547	8.5	849	US-09-252-991A-17953	Sequence 17953, A
17	545	8.5	106	US-09-463-702A-43	Sequence 43, Appl1
18	545	8.5	106	US-09-699-135-43	Sequence 43, Appl1
19	529	8.2	616	US-09-107-532A-7258	Sequence 7258, Ap
20	526	8.2	637	US-09-134-001C-5658	Sequence 5658, Ap
21	525.5	8.2	1328	US-08-781-891-76	Sequence 76, Appl1
22	525.5	8.2	1328	US-09-618-166-76	Sequence 76, Appl1
23	521	8.1	473	US-09-107-532A-5941	Sequence 5941, Ap
24	506.5	7.9	410	US-09-949-016-9327	Sequence 9327, Ap
25	494	7.7	355	US-09-902-540-15688	Sequence 15688, A
26	487	7.6	393	US-08-559-303B-76	Sequence 76, Appl1
27	487	7.6	393	US-09-175-828-76	Sequence 76, Appl1

28	487	7.6	393	US-09-753-143-76	Sequence 76, Appl1
29	485.5	7.6	1401	US-09-127-670-6	Sequence 6, Appl1
30	481.5	7.5	333	US-09-463-702A-39	Sequence 39, Appl1
31	481.5	7.5	333	US-09-699-135-39	Sequence 39, Appl1
32	479.5	7.5	1401	US-08-781-891-206	Sequence 206, App
33	479.5	7.5	1401	US-09-618-166-206	Sequence 206, App
34	478.5	7.4	330	US-09-134-000C-4772	Sequence 4772, Ap
35	476	7.4	1409	US-09-949-001-22	Sequence 22, Appl1
36	475	7.4	1269	US-08-781-891-74	Sequence 74, Appl1
37	475	7.4	1269	US-09-618-166-74	Sequence 74, Appl1
38	475	7.4	1432	US-08-781-891-71	Sequence 71, Appl1
39	475	7.4	1432	US-09-618-166-71	Sequence 71, Appl1
40	475	7.4	1432	US-09-949-001-16	Sequence 16, Appl1
41	473	7.4	879	US-09-248-796A-20040	Sequence 20040, A
42	464.5	7.2	386	US-08-559-303B-75	Sequence 75, Appl1
43	464.5	7.2	386	US-09-175-828-75	Sequence 75, Appl1
44	464.5	7.2	386	US-09-753-143-75	Sequence 75, Appl1
45	464.5	7.2	659	US-08-781-891-77	Sequence 77, Appl1

ALIGNMENTS

RESULT 1

US-09-463-702A-2

Sequence 2, Application US/09463702A

Patent No. 635435

GENERAL INFORMATION:

APPLICANT: AGENE Research Institute, Co., Ltd.

APPLICANT: HIRAKI AND ASSOCIATES

APPLICANT: SHIMAMOTO, AKIRO

APPLICANT: KITAO, SAORI

APPLICANT: FURUICHI, YASUHIRO

TITLE OF INVENTION: HUMAN GENE RECD4 ENCODING HELICASE

FILE REFERENCE: HIRALI50

CURRENT APPLICATION NUMBER: US/09/463,702A

CURRENT FILING DATE: 2000-01-24

PRIOR APPLICATION NUMBER: PCT/JP98/03114

PRIOR FILING DATE: 1998-07-10

PRIOR APPLICATION NUMBER: JAPAN 9/200387

PRIOR FILING DATE: 1997-07-25

NUMBER OF SEQ ID NOS: 44

SOFTWARE: PatentIn version 3.0

SEQ ID NO 2

LENGTH: 1208

TYPE: PRT

ORGANISM: Homo sapiens

US-09-463-702A-2

Query Match

Best Local Similarity 100.0%; Score 6424; DB 3; Length 1208;

Matches 1208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MERLDVRRRLQAWERAFRRPSQDDVEAABETRLYREYTLKRTTGOAGGUR	60
DB	1	MERLDVRRRLQAWERAFRRPSQDDVEAABETRLYREYTLKRTTGOAGGUR	60
QY	61	SEESLPAAAEAPPCPCPHINRAATKSPOPTPGSRGSGYPDQGRKANUKTGLQNG	120
DB	61	SEESLPAAAEAPPCPCPHINRAATKSPOPTPGSRGSGYPDQGRKANUKTGLQNG	120
QY	121	PALGRPPWPLGASASASTPKPCGTPVSPFAEKVSDEBPOLPEPQPRGRLOHLQASIS	180
DB	121	PALGRPPWPLGASASASTPKPCGTPVSPFAEKVSDEBPOLPEPQPRGRLOHLQASIS	180
QY	181	QRLGSLDPGWLORCHSEVDFLGAAPACRDPDGSBESQULLIGESAIVLPGAGSQGPBAS	240
DB	181	QRLGSLDPGWLORCHSEVDFLGAAPACRDPDGSBESQULLIGESAIVLPGAGSQGPBAS	240
QY	241	ARQEVSIKVGSGQPPSSGGEKRRNNEEPWESPAQVQDSSQAGPPBEGAGVAVEEDPPG	300
DB	241	ARQEVSIKVGSGQPPSSGGEKRRNNEEPWESPAQVQDSSQAGPPBEGAGVAVEEDPPG	300

QY	301	BPVQOPPOPCSSPENPNHYHGISSESSQABAGAEETAPLHTEPRRLAHDHGRVYVLANMKQ	360
Db	301	BPVQOPPOPCSSPENPNHYHGISSESSQABAGAEETAPLHTEPRRLAHDHGRVYVLANMKQ	360
QY	361	KHYVGRALRSLLKQAMKQKMRKKGEFCFGGGAIVTTTKESCFLNEQFDHMAACPPRA	420
Db	361	KHYVGRALRSLLKQAMKQKMRKKGEFCFGGGAIVTTTKESCFLNEQFDHMAACPPRA	420
QY	421	SEEDTAVGPEPLVSPQPVPEVPSLDPTVLPLYSIGPSGQALETPAEVFOALEQGHQA	480
Db	421	SEEDTAVGPEPLVSPQPVPEVPSLDPTVLPLYSIGPSGQALETPAEVFOALEQGHQA	480
QY	481	FRPGGERAMVRLSGISTLLVLP.TGAGSLCYQALYSRSPCLTVVSVLSLMDQ	540
Db	481	FRPGGERAMVRLSGISTLLVLP.TGAGSLCYQALYSRSPCLTVVSVLSLMDQ	540
QY	541	VSGLEPPLCKAACIHSGMTRKQRESEYLOKIRAAQVHMLTPPALVAGAGLPPAOLPVA	600
Db	541	VSGLEPPLCKAACIHSGMTRKQRESEYLOKIRAAQVHMLTPPALVAGAGLPPAOLPVA	600
QY	601	FACIDBAHCLSQWNSHNPFCYLAVCKVLNERMGVHCFGLTATRTRTASDVAOHLVAE	660
Db	601	FACIDBAHCLSQWNSHNPFCYLAVCKVLNERMGVHCFGLTATRTRTASDVAOHLVAE	660
QY	661	BEDHNGPAPVPNNLH.SVSMEDPTOALLTLLOGRFOWLDSIIYYCNRREDTEI.AALL	720
Db	661	BEDHNGPAPVPNNLH.SVSMEDPTOALLTLLOGRFOWLDSIIYYCNRREDTEI.AALL	720
QY	721	RTCLAAAWPGSGGRAPKTTAAEYHAGMCSRERRRQORA.FMFGOLRVVVA.TYAFMGGLDR	780
Db	721	RTCLAAAWPGSGGRAPKTTAAEYHAGMCSRERRRQORA.FMFGOLRVVVA.TYAFMGGLDR	780
QY	781	PDVRAVLHLGLRPSFESYVOAVGARGDQPAHCH.FLOPQGEDLRELRHHVADSTFL	840
Db	781	PDVRAVLHLGLRPSFESYVOAVGARGDQPAHCH.FLOPQGEDLRELRHHVADSTFL	840
QY	841	AVKVLVORVFPACTCTCTPRSEBGAVGGERPVKPYRQEA.FOJLSHOABPBRKVCMH	900
Db	841	AVKVLVORVFPACTCTCTPRSEBGAVGGERPVKPYRQEA.FOJLSHOABPBRKVCMH	900
QY	901	ERALPIOTVOALDMPBEAIEITLLCYLELHPHMYELATTYTHCR.LNPGGPAOLOALA	960
Db	901	ERALPIOTVOALDMPBEAIEITLLCYLELHPHMYELATTYTHCR.LNPGGPAOLOALA	960
QY	961	HRCPLAVCLAOQLBEDPCQSSSVYEDVWXLVDSMGWELASVRRAL.COLOWDHEBRTGV	1020
Db	961	HRCPLAVCLAOQLBEDPCQSSSVYEDVWXLVDSMGWELASVRRAL.COLOWDHEBRTGV	1020
QY	1021	RRGTVLVYFESLAFHLNPSFGULTABEKQICDPL.YGRVOAERQALARIPTFOAFHSV	1080
Db	1021	RRGTVLVYFESLAFHLNPSFGULTABEKQICDPL.YGRVOAERQALARIPTFOAFHSV	1080
QY	1081	AFPSGCPCLIEQODEERSYTRLXDLGRYFEESGOEPGCMEDAQPEPGGARLODMEDEVK	1140
Db	1081	AFPSGCPCLIEQODEERSYTRLXDLGRYFEESGOEPGCMEDAQPEPGGARLODMEDEVK	1140
QY	1141	CDIROFLSLRPEKESSRAVARI.FHGIGSPCYPAOYGOGRRFWRKXYLHLSFHALVGLAT	1200
Db	1141	CDIROFLSLRPEKESSRAVARI.FHGIGSPCYPAOYGOGRRFWRKXYLHLSFHALVGLAT	1200
QY	1201	BELLQVAR 1208	
Db	1201	BELLQVAR 1208	
RESULT 2			
US-09-699-135-2			
; Sequence 2, Application US/09699135			
; Patent No. 6472513			
; GENERAL INFORMATION:			
; APPLICANT: AGENE Research Institute, Co., Ltd.			
; APPLICANT: HIRAKI AND ASSOCIATES			
; APPLICANT: SHIMAMOTO, AKIRO			

	APPLICANT: KITAO, SAORI	APPLICANT: FURUTACHI, YASUHIRO	TITLE OF INVENTION: HUMAN GENE SEQ4 ENCODING HELICASE	FILE REFERENCE: HIRAI150	CURRENT APPLICATION NUMBER: US/09/699,135	CURRENT FILING DATE: 2000-10-27	PRIOR APPLICATION NUMBER: US/09/463,702A	PRIOR FILING DATE: 2000-01-24	PRIOR APPLICATION NUMBER: PCT/Jp98/03114	PRIOR FILING DATE: 1998-07-10	PRIOR APPLICATION NUMBER: JAPAN 9/200387	PRIOR FILING DATE: 1997-07-25	NUMBER OF SEQ ID NOS: 44	SOFTWARE: PatentIn version 3.0	SEQ ID NO 2	LENGTH: 1208	TYPE: PRT	ORGANISM: Homo sapiens	US-09-699-135-2
QY	Query Match	100.0%; Score 6424; DB 4; Length 1208;	Best Local Similarity 100.0%; Pred. No. 0;	Matches 1208; Conservative 0; Mismatches 0; Indels 0; Gaps 0															
QY	1	MERLADVVERIQAWRARPRQRGRPPSDDDVEAAPEERFALYREYRTLLKRTTGOAGGGR	60																
DB	1	MERLADVVERIQAWRARPRQRGRPPSDDDVEAAPEERFALYREYRTLLKRTTGOAGGGR	60																
QY	61	SSESIPAAAEAPBERCWPHLNRAATKSPQPTPGRSQGSVPDYGQRLKANIKGTLQAG	120																
DB	61	SSESIPAAAEAPBERCWPHLNRAATKSPQPTPGRSQGSVPDYGQRLKANIKGTLQAG	120																
QY	121	PALGRPWPPLGRASSKASTPKPCTGVPVSPAKEYSDPPQLPEPQPRPGRLOHLQASIS	180																
DB	121	PALGRPWPPLGRASSKASTPKPCTGVPVSPAKEYSDPPQLPEPQPRPGRLOHLQASIS	180																
QY	181	QRLGSLDPGCMQORCHSEVPDPLGAKACRPRDLGSEBSQLITGSAVILGPGAGSQGPES	240																
DB	181	QRLGSLDPGCMQORCHSEVPDPLGAKACRPRDLGSEBSQLITGSAVILGPGAGSQGPES	240																
QY	241	AFQEVSIWGSPOPSSSGGCKRNMEEPMWESPAAQVOESQAGPSEAGAAVAVEDPPG	300																
DB	241	AFQEVSIWGSPOPSSSGGCKRNMEEPMWESPAAQVOESQAGPSEAGAAVAVEDPPG	300																
QY	301	EPVQAQPPQPCSSPSPNPRYHGLSPSSQAPAGKAEGTAPLPIFPRLARHDRGNVYRLNMKQ	360																
DB	301	EPVQAQPPQPCSSPSPNPRYHGLSPSSQAPAGKAEGTAPLPIFPRLARHDRGNVYRLNMKQ	360																
QY	361	KHYVGRALRSRLRKQAKQMKRKKGCFGGGAGTYYTKSCFPLNEQPDHMAAOCPPRA	420																
DB	361	KHYVGRALRSRLRKQAKQMKRKKGCFGGGAGTYYTKSCFPLNEQPDHMAAOCPPRA	420																
QY	421	SEEDTDAGPEPLVPSPOQVPEVPSLDPVLPLYSLSFGSQLAETPAEYFQALQEOIGHQ	480																
DB	421	SEEDTDAGPEPLVPSPOQVPEVPSLDPVLPLYSLSFGSQLAETPAEYFQALQEOIGHQ	480																
QY	481	FRPGQERAVMRILSGISITLLVLPFGAGKSLCYQLPALLYSRRSPCLTLVVSPLSLMDQ	540																
DB	481	FRPGQERAVMRILSGISITLLVLPFGAGKSLCYQLPALLYSRRSPCLTLVVSPLSLMDQ	540																
QY	541	VSGLPBPCLKAACHSGMTRKQRESVLQKIRAAQVHVMLTTPALVAGAGLPAAQLPPVA	600																
DB	541	VSGLPBPCLKAACHSGMTRKQRESVLQKIRAAQVHVMLTTPALVAGAGLPAAQLPPVA	600																
QY	601	FACIDAEACLQSWNSNFRPCYLIRVCKVLREBMGVHCFGLTATARTASDVAAQHLAAVE	660																
DB	601	FACIDAEACLQSWNSNFRPCYLIRVCKVLREBMGVHCFGLTATARTASDVAAQHLAAVE	660																
QY	661	EPDLHGAPVPVFNLIHVSMDRDTQALLTLQGRFQNLDSIIITYCNREDTERIALLL	720																
DB	661	EPDLHGAPVPVFNLIHVSMDRDTQALLTLQGRFQNLDSIIITYCNREDTERIALLL	720																
QY	721	RTCHAAVAVPGSGGAPPTTAAEAYHAGMCSRRRRVQAPMGOQLRVVATVAFGKGLDR	780																

DB 721 RTCHAAVPPSGGAPKTTAAVYHAGNCSRRRRVQRAFMGQLRVVATAVAFMGGLDR 780
QY 781 PDVRAVHLGLPSPESYVQAVGRAGDQPAHCHFLQPOGEDI,RELRRVHADSTDFL 840
DB 781 PDVRAVHLGLPSPESYVQAVGRAGDQPAHCHFLQPOGEDI,RELRRVHADSTDFL 840
QY 841 AVKRLVORVFACTCTCRPPSEOGAVGGERPVKPYRPOEABOLSHOAPGPRVCMKH 900
DB 841 AVKRLVORVFACTCTCRPPSEOGAVGGERPVKPYRPOEABOLSHOAPGPRVCMKH 900
QY 901 ERALPQLTVQALDMPBEALETILCYLHPPHMLLATTYTHCLNCPGPAQLQALA 960
DB 901 ERALPQLTVQALDMPBEALETILCYLHPPHMLLATTYTHCLNCPGPAQLQALA 960
QY 961 HRCPLAVCLAQOLPEDPGOSSSVFEDWKLVDMSGWEIASVRALCOLQMDHEPRGV 1020
DB 961 HRCPLAVCLAQOLPEDPGOSSSVFEDWKLVDMSGWEIASVRALCOLQMDHEPRGV 1020
QY 1021 RRGTVLVFBSLAFHLSRGLTAEKQICDPLVGRVQAREQALARTFOAFHSV 1080
DB 1021 RRGTVLVFBSLAFHLSRGLTAEKQICDPLVGRVQAREQALARTFOAFHSV 1080
QY 1081 AFPSCGPLEQODEBERSTRKDLGRYFEEBEGEBGEMDAQEPGQARLQDMEQYR 1140
DB 1081 AFPSCGPLEQODEBERSTRKDLGRYFEEBEGEBGEMDAQEPGQARLQDMEQYR 1140
QY 1141 CDIROPFLRPEKFSRAVARIFHIGISPCYPAQVYQDRFRFRKYLHLSFHALVGLAT 1200
DB 1141 CDIROPFLRPEKFSRAVARIFHIGISPCYPAQVYQDRFRFRKYLHLSFHALVGLAT 1200
QY 1201 BELLQVAR 1208
DB 1201 BELLQVAR 1208

RESULT 3

US-09-463-702A-37
; Sequence 37, Application US/09463702A
; Patent No. 6335435
; GENERAL INFORMATION:
; APPLICANT: AGENE Research Institute, Co., Ltd.
; APPLICANT: HIRAKI AND ASSOCIATES
; APPLICANT: SHIMAMOTO AKIRO
; APPLICANT: KITNO, SAORI
; APPLICANT: FURUICHI, YASUHIRO
; TITLE OF INVENTION: HUMAN GENE RECO4 ENCODING HELICASE
; FILE REFERENCE: HIRAI150
; CURRENT APPLICATION NUMBER: US/09/463,702A
; PRIOR FILING DATE: 2000-01-24
; PRIOR APPLICATION NUMBER: PCT/JP98/03114
; PRIOR FILING DATE: 1998-07-10
; PRIOR APPLICATION NUMBER: JAPAN 9/200387
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 37
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-463-702A-37

Query Match 29.3%; Score 1883; DB 3; Length 361;
Best Local Similarity 100.0%; Pred. No. 3.7e-144;
Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 471 QALBOLGHOAFRPPGERAVMRILSGISTLLVPTGAGSKLCYQLPALYSRRSPCLTLV 530
DB 1 QALBOLGHOAFRPPGERAVMRILSGISTLLVPTGAGSKLCYQLPALYSRRSPCLTLV 60
QY 531 SPLSLMDQVSGLPPLCKAACISGMRKORESVLQKIRAAOVHVLMLTPBALVGAGL 590
DB 61 SPLSLMDQVSGLPPLCKAACISGMRKORESVLQKIRAAOVHVLMLTPBALVGAGL 120

QY 591 PPAOLPVAFACTIDEAHCISQMSHNPBCYLKVCULREBVGVCFLGTATATRTAS 650
DB 121 PPAOLPVAFACTIDEAHCISQMSHNPBCYLKVCULREBVGVCFLGTATATRTAS 180
QY 651 DVAOHLAAVEBBDLGPAPVPTNLHLSVSMRDTQALTLTLQGRFONLDSIIYYCNR 710
DB 181 DVAOHLAAVEBBDLGPAPVPTNLHLSVSMRDTQALTLTLQGRFONLDSIIYYCNR 240
QY 711 EDTERIALLRTCLHAAVPPSGGAPKTTAAVYHAGNCSRRRRVQRAFMGQLRVVVA 770
DB 241 EDTERIALLRTCLHAAVPPSGGAPKTTAAVYHAGNCSRRRRVQRAFMGQLRVVVA 300
QY 771 TVAFMGGLDRPDVRAVHLGLPSPESYVQAVGRAGDQPAHCHFLQPOGEDI,RELRR 830
DB 301 TVAFMGGLDRPDVRAVHLGLPSPESYVQAVGRAGDQPAHCHFLQPOGEDI,RELRR 360
QY 831 H 831
DB 361 H 361

RESULT 4

US-09-699-135-37
; Sequence 37, Application US/09699135
; Patent No. 6472513
; GENERAL INFORMATION:
; APPLICANT: AGENE Research Institute, Co., Ltd.
; APPLICANT: HIRAKI AND ASSOCIATES
; APPLICANT: SHIMAMOTO AKIRO
; APPLICANT: KITNO, SAORI
; APPLICANT: FURUICHI, YASUHIRO
; TITLE OF INVENTION: HUMAN GENE RECO4 ENCODING HELICASE
; FILE REFERENCE: HIRAI150
; CURRENT APPLICATION NUMBER: US/09/699,135
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: US/09/463,702A
; PRIOR FILING DATE: 2000-01-24
; PRIOR APPLICATION NUMBER: PCT/JP98/03114
; PRIOR FILING DATE: 1998-07-10
; PRIOR APPLICATION NUMBER: JAPAN 9/200387
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 37
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-699-135-37

Query Match 29.3%; Score 1883; DB 4; Length 361;
Best Local Similarity 100.0%; Pred. No. 3.7e-144;
Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 471 QALBOLGHOAFRPPGERAVMRILSGISTLLVPTGAGSKLCYQLPALYSRRSPCLTLV 530
DB 1 QALBOLGHOAFRPPGERAVMRILSGISTLLVPTGAGSKLCYQLPALYSRRSPCLTLV 60
QY 531 SPLSLMDQVSGLPPLCKAACISGMRKORESVLQKIRAAOVHVLMLTPBALVGAGL 590
DB 61 SPLSLMDQVSGLPPLCKAACISGMRKORESVLQKIRAAOVHVLMLTPBALVGAGL 120
QY 591 PPAOLPVAFACTIDEAHCISQMSHNPBCYLKVCULREBVGVCFLGTATATRTAS 650
DB 121 PPAOLPVAFACTIDEAHCISQMSHNPBCYLKVCULREBVGVCFLGTATATRTAS 180
QY 651 DVAOHLAAVEBBDLGPAPVPTNLHLSVSMRDTQALTLTLQGRFONLDSIIYYCNR 710
DB 181 DVAOHLAAVEBBDLGPAPVPTNLHLSVSMRDTQALTLTLQGRFONLDSIIYYCNR 240
QY 711 EDTERIALLRTCLHAAVPPSGGAPKTTAAVYHAGNCSRRRRVQRAFMGQLRVVVA 770
DB 241 EDTERIALLRTCLHAAVPPSGGAPKTTAAVYHAGNCSRRRRVQRAFMGQLRVVVA 300


```

Db 74 ISLMDQVQLQANGVAACLNSGTREQLQEVMTGCRGTQIRLLIYIAPERLMLDNFLSH 133
QY 593 AAQLPVPACIDEAHCLSSQSHNRPCLRVCKYLREMGVHCLGLTATRTPTASDV 652
Db 134 LAHNPVLLA-VDEAHCLSSQSHNRPCLRVCKYLREMGVHCLGLTATRTPTASDV 191
QY 653 AQLHVAEEDLHGAPVPTNLHLSVSMRDTQALTLTGKRFONLDSI----- 703
Db 132 VRLLG-----LNDPL-----IQIS-SFDRPNIRYMLM-----EKFPDQLMRVYQEQG 235
QY 704 ---IYCNREDTERIALLRCLHAANVPGSGGAPKTTAAYAHGMSRERRRVQRAF 760
Db 236 KSGIITCNSRAKVEDTAAAL-----QSKGISAAAYHAGLENNVRADVQEKF 281
QY 761 MGGQLRVVAVATVAFMGGLDRPDVRAVLHLGLPPSFESYVQAVGRAGDQPAHCHLFL-- 818
Db 282 QRDQLQIVAVATVAFMGINKPVRVHFDPINIESYYQETGRAGDGLPAEAMLFYDP 341
QY 819 -----QPGEDLRELRH---VHADSTDFLAVKRLV-----QVFPACTG- 855
Db 342 ADMAWLRCLREKPPQGG-LQDIERHKLAMGAPAEAGTCRRLVLTNYFEGRGQEPGNCND 400
QY 856 TCTRPPEQEGAVGE 871
Db 401 ICLDPPKQYDGSSTDAQ 416

```

RESULT 7
US-08-559-303B-77
Sequence 77, Application US/08559303B
Patent No. 5824501
GENERAL INFORMATION:
APPLICANT: NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA
APPLICANT: GRODEN
TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT
TITLE OF INVENTION: OF BLOOM'S SYNDROME
NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
STREET: 90 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/559,303B
FILING DATE: NOVEMBER 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: ELIZABETH A. BOGOSIAN
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 63475/65
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
INFORMATION FOR SEQ ID NO: 77:
SEQUENCE CHARACTERISTICS:
LENGTH: 370
TYPE: AMINO ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: OTHER NUCLEIC ACID
ANTHETICAL: YES
ANTI-SENSE: NO
FEATURE:
NAME/KEY:
LOCATION:

```

IDENTIFICATION METHOD:
OTHER INFORMATION:
US-08-559-303B-77
Query Match 9.7%; Score 621; DB 2; Length 370;
Best Local Similarity 40.0%; Pred. No. 6,7e-42;
Matches 154; Conservative 57; Mismatches 110; Indels 64; Gaps 12;

```

474 BGLGQADRPQGERAVMBILIGISTLVLPNGAGKSLCYOLPALLYSRRSPCLTVSPSL 533
4 EYFGYQFRPGQEBITDVLGSRDCLVMPFGGKSLCYQIPALLING---LTVVSPSL 59
534 ISLMDQVSGLPP-CLKACIHSGMTKQRBESVLQKIRPAQVHVMTPEALVAGGLPP 592
60 ISLMDQVQLQANGVAACLNSGTREQLQEVMTGCRGTQIRLLIYIAPERLMLDNFLSH 119
593 AAQLPVPACIDEAHCLSSQSHNRPCLRVCKYLREMGVHCLGLTATRTPTASDV 652
120 LAHNPVLLA-VDEAHCLSSQSHNRPCLRVCKYLREMGVHCLGLTATRTPTASDV 177
653 AQLHVAEEDLHGAPVPTNLHLSVSMRDTQALTLTGKRFONLDSI----- 703
178 VRLLG-----LNDPL-----IQIS-SFDRPNIRYMLM-----EKFPDQLMRVYQEQG 221
704 ---IYCNREDTERIALLRCLHAANVPGSGGAPKTTAAYAHGMSRERRRVQRAF 760
222 KSGIITCNSRAKVEDTAAAL-----QSKGISAAAYHAGLENNVRADVQEKF 267
761 MGGQLRVVAVATVAFMGGLDRPDVRAVLHLGLPPSFESYVQAVGRAGDQPAHCHLFL-- 818
268 QRDQLQIVAVATVAFMGINKPVRVHFDPINIESYYQETGRAGDGLPAEAMLFYDP 327
819 -----QPGEDLRELRH 831
328 ADMAWLRCLREKPPQGG-LQDIERH 351

RESULT 8
US-09-175-828-77
Sequence 77, Application US/09175828
Patent No. 6221643
GENERAL INFORMATION:
APPLICANT: NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA
APPLICANT: GRODEN
TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT
TITLE OF INVENTION: OF BLOOM'S SYNDROME
NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
STREET: 90 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/175,828
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/559,303
FILING DATE: NOVEMBER 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: ELIZABETH A. BOGOSIAN
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 63475/65
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766


```

QY      ACHLAABEERPHADPAVPYTMILHSVSMEDRTODATLTLQKFFOMJDSI----- 703
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      VRLIG-----LNPPL-----IQVS-SFDRNIRMYLM-----EKFKPLDQMLRYYVDORG 247
QY      704 ---IYYCNRREDTERIALALFTCLHAAMPVSGSGAPAXTAAEYHAGCNSERRRORAF 760
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      KSGIITICNSRSKVEDTARLQ-----SRG-----ISAAAYHAGLENNVRAVEOKEF 293
QY      MOGQLRVVVAVVAFGMDLRDPDVAVLHLGLPSPESYVQAVGRAGRGOPAHCHLFLQRP 820
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      QRDQLQIVVAIVAGMGINKENVRVVFHDI.PRNIESIYQGTGARGDGLPAEAMLFDP 353
QY      821 -----QGEDIRELRRH---VHADSTDFLAVKLV-----QRFVPACTC-T 856
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      ADMAMLRCLCEBKPAQFLQDIERRHKLNAMGAFAAQTCRRRLVLTLYNFEGRQEGCNCIDI 413
QY      CTRPSPBEGEAVGGERVYPK-YPPQEAQLSH-----QAAGPRRVCMCHGRALPI 906
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      CLDPBPQDGLMDARKALSTYIRVNRQRFPMGVVVEVLKGANNOIRRENGHK-LPV 468

```

```

RESULT 13
US-09-543-681A-4359
; Sequence 4359, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709,1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4359
; LENGTH: 641
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-4359

```

```

Query March 9.2%: Score 591.5; DB 4; Length 641;
Best Local Similarity 31.9%; Pred. No.3.7e-39;
Matches 187; Conservative 71; Mismatches 153; Indels 175; Gaps 25;

Qy 467 AEVFQAL-----EQLGQAFRRGGDERAVMRILSGISTLLVLPYGKSLCYQLPALI 518
Db AEVNTMSAQVILRETRGTGYQFRGGQOEIIDTIIITGRDCLVMPPTGGKSLCYQIPALL 36
Qy 519 YSRSPCLTLYVSPILSLMDPOVSGLPCL--KAACIHSGMTRKQRESVYLQKIRAAOVH 575
Db 96 LDG---LTVVSPILSLMKQVQDOL--CLHGIDAAFAFNSTQSDQEQVQVMRCQRGIRK 149
Qy 576 VLMTPREALVAGGGLPPAAQLPVAFACIDBAHCISQWNSHFRPCYLRVCYKILERMGVH 635
Db 150 LLTYAPRLMWSFLHHLVQWQF--ALLAVDEAHCSQMGHPFREY--RGILGLAQYLPDV 207
Qy 636 CFLGLTATATRTKTSVDVAQHVAVAEEDLDGPAVPYTNLHLSVSMRDPDQALLTLIQK 695
Db 208 PIALTAADNTTRHDIIINOLALR-----TPV---VHIS-SFDRRNIRYTLV---E 251
Qy 696 RFQNDLSI-----IYCNREDTERIALALRTCLHAAMVPSGGRAPIKTTAE 743
Db 252 KYKPLDQMLFTRGQKGSGLIYCNRSKSVETAEFL-----GKRG---LSIAA 297
Qy 744 YHAGCSEERRRVRQAFMQQLRVVATVAFAQMGIDRDPVAVLHGLPSPFESYVQAVG 803
Db 298 YHAGMDITRAKVQDAFQRDQLQIVATVAFAQMGINKENVFVYFDIPRNIESYQGTG 357
Qy 804 RAGSDGQGAHCHLFLQPGEDLRRLRHVHADSTDFLAVKVLQVRVFPACTCTCRPSE 863
Db 358 RAGSDGRLPAQAVLFDLP--ADMAWMLR-----CLBEKPSE 391

```

```

Oy      864  QEAGVGEPRVPKYPPOEAEOLSHQAAAGPRVVCNGHERAPLTQTLADMEAEIETL  923
          :::::
Db      392  QKAI-----BSHKLNAMGAFAEAQTC-----RL-----VL  417
          :::::
Oy      924  LCYLELPHPHMLELTATYTHCRILNCPGAPALQALAHRCPELVACLAQQLPEDPGGSS  983
          :::::
Db      418  LNTF-----GHRQQA-----GNCNICTL-----DPRK---  440
          :::::
Oy      984  SVEFDNWKLVDSNGWELASVRPALCOLQMDHBPRTGVRRTGVAVE  1029
          :::::
Db      441  --GYD--GLVDA-----QKALACIY-----RIGQRFGIGYIVE  469

```

```

RESULT 14
US-09-902-540-16497
; Sequence 16497, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(1584)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ. ID NOS: 16825
; SEQ ID NO 16497
; LENGTH: 1394
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-16497

```

Query Match	9.1%	Score 585	DB 4	Length 1394
Best Local Similarity	26.8%	Pred. No. 3.9e-38		
Matches 261	Conservative 118	Mismatches 347	Indels 248	Gaps 44
QY	2	ERLRDV--RERLQW--ERAFR-----	-RQGRPRQDDVEAPAEETRALYREYRLTK	49
DB	240	QRVVDLYCKRLQAMHEDHWKRVTTVITAAVTSAGAPRVYDRQSGTQVEQ--GKVLVD	297	
QY	50	RTTQOAGGGLSSSSLPAAAEAP--EPRCNGPHLNRAATYSPQPTPGSRGQSVDDYQO	107	
DB	298	IGGGQKAPRLKSEGRKGEDAREPEDEPQTLPSGLARAQPSVEDVEAWKKTTRPP--R	355	
QY	108	RLKANLKTQLQ-AGPAL-----GRRWPLGRASSKA-----	ST 139	
DB	356	FTDATLTLLTAMASAGRALDEKELADAMRTGTGTPTATRAIIEVLDDREYLRRRKVMET	415	
QY	140	PK-----PPTGPVPVSEFAKVSDEPQLPEPCRPRLQHLQASLSQRLGS	185	
DB	416	EKGHILIOVNPDKYTPAMTQGMAMLQRIERGEQDLB-----FNGIEAYIVLEVGH	469	
QY	186	LDPGMLQCHSEVPDFLQAPKACRPDLQ--SEBSQLLIPGESAVLGPAGSQGBEA-SA	241	
DB	470	GATAPQRPFRPEV---LGAP---PGAGMGQGAQAQVASEGAAAHAGAGARGGCGQA	522	
QY	242	FOEYVSI-RVSGPPOSSSGGGRKRNNEEWPSPAOVQOQSSSQAGPPSEBAGAVAVEDPPG	300	
DB	523	FAESSAEERFGD--ASAGAGARGGAGATEASA--ERFPAALAPGVQOGR-----	PG 569	
QY	301	EPVQAQRPQSPSPSNPRYHGLSPSSQARAGALEGTAPLHIFPRLARHDRGNVYRLNMKQ	360	
DB	570	ALSESESP-----LFGTRKRRTSGSVRSERPH-----GEPPLSTWSP	606	
QY	361	KHYVGRALRSRLLRKQ--AWKQ--KWRKKGEFGGGGATVTTKSCFLNEQFDHMAACP	417	
DB	607	-----GGGAARLPLPADGTGMNAGAMENTG-----AAAP	636	
QY	418	-----RPASEEDTDAVGPPELVPSQVPEVPSLDTPLYLPLYSLCPSCQLAETTPAEVFOA	472	

```

Db      637  VNRARTPPASASTG-----GRPERVARRAPTAPALRFLK-----671
Qy      473  LEOLGHQARPCQOEAWRILSGISITLVLPFGAGSLCYQLPALYSRRSPCLTLVSP 532
Db      672  -EAFSPDRPYOEAVCRATAGBEDLLVMPGAGKSLCYQLPGLRAG-----TLVSP 726
Qy      533  LILNDPOVSGLPKC-LKACIHSQWTRKQRESVLQKIRAAQVHVMLELPELVAGGAP 591
Db      727  LIALMEDQVARIQSIGFADRIHSGDRMSRQVCADYIEDRIDFLFIAPERLGVPGFYE 786
Qy      532  PAAQLPVAFCIDEACISQWSHNRPCYLKVCYLKRMGV---HCFIAGTATART 648
Db      787  FLARTPT-ALIVADENHCISQWGHDRPY-----RLGARLPYLRRAPVALTATITPV 841
Qy      649  ASDVAQHLVAABEEDLHGAPVP-----TNLHLSVSMDDTD---QALLTLQ 694
Db      842  QRDIVQQLG-----LQPGGKARFTFHGFRRTNIAIV---RELNPGARGDALQGLLED 892
Qy      695  KRFOULDSIIICNRREDTERIAALLRTCLHAAWVSGSGRAPKTTAEVHAGMCSRRER 754
Db      893  EENR---PAIVAAATRKAAEQULADOL-----AGEPP---AAAYHAGLOPSEED 934
Qy      755  RVQAFMGOQLRVVAVTVAFGMLDRPDYRAVLHLGLPPSPESYQAVAGRAGDQPAHC 814
Db      935  RVQAFELRGLSLVIVATTAFGMDKADVRVTHALPRLASLGYYQELGRAGRDGKPSRA 994
Qy      815  --HLFLOPQEDLRELRRHVDSTDFLAVKRLVQVFPACTCTCTPPSBOGAVGE 871
Db      995  VLLHSYIDRRTHEFFRRDYPEA-----VLERLFKS-TAPQLEPKAVLQGRVGD 1044
Qy      872  RVPKYPPEQAEOL 885
Db      1045  ---PEVFDALAEOL 1055

```

```

RESULT 15
US-09-134-000C-5710
; Sequence 5710, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lytn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: Patent version 3.1
; SEQ ID NO 5710
; LENGTH: 589
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5710

```

```

Query Match      8.6%; Score 554.5; DB 4; Length 589;
Best Local Similarity 34.0%; Pred. No. 3,2e-36;
Matches 143; Conservative 67; Mismatches 147; Indels 63; Gaps 10;

```

```

Qy      472  ALEOL-----GHQARPCQOEAWRILSGISITLVLPFGAGSLCYQLPALYSRRSPCL 526
Db      3  ALQELLLKTFGYDDRPQGETIIRHVRQENVLGIMPFGGKSLCYQLPALLDN---L 58
Qy      527  TLVSPPLSLMDQVS-----GLPPLKACIHSQWTRKQRESVLQKIRAAQVHVMLEP 581
Db      59  TLVISPPLSLMDQVDALNMGIP-----ATYINSTISYQENMHRITQLANKEVKLLVAP 114
Qy      582  EALVAGGLPRAQLPVAFCIDEACISQWSHNRPCYLKVCYLKRMGVHCFGLT 641
Db      115  ERLBSYDFQOQMLTHV-PLDLLAVDEAHCSQWGHDRFPSTRLAETIIDQFOOQPTVIALT 173

```

```

Qy      642  ATATRTASDVAAQHLVAABEEDLHGAPVPNTLHLSVSMDDTDQALLTLQGRFONLD 701
Db      174  ATATPOVAEDIVKQLRIPSENEIK-TGFARENLSFOVKDQNRDVPFLFIYLMNTGQ--- 229
Qy      702  SIIICNRREDTERIAALLRTCLHAAWVPGSGRAPKTTAEVHAGMCSRRERRVQORAFM 761
Db      230  SGIIVASTRKEVERIYHLL-----ESKTLAAGMYHGMSPQLRSENQEAFL 275
Qy      762  QGQLRVVAVTVAFGMLDRPDYRAVLHLGLPPSPESYQAVAGRAGDQPAHCILFLOPQ 821
Db      276  YDQVQVMAVTAFAFGMINKSNVRFVTHQVPGNIESYQOAGRAGRDGLPSDAVIMFAPO 335
Qy      822  GED-----LRELRRHVDSTDFLAVKRLVQVFPACTCTCTR 859
Db      336  DLQIQYFLBOSEMTIDYKQKEYLKLREMSQYAAHQ---KCLQXYILRYFGEETDCGR 391

```

```

Search completed: March 19, 2005, 01:08:46
Job time : 49 secs

```


CC useful in the diagnosis, especially prenatal diagnosis, and treatment of
CC Rothmund-Thomson syndrome. The genomic sequence may especially be used in
CC gene therapy for this condition

xx Sequence 1208 AA;

Query Match 100.0%; Score 6424; DB 3; Length 1208;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MERLDVBERLOAMRARRRQGRPPSODDVEAAEEETRALYREYRTTLKRTTGQGGGIR 60
DB 1 MERLDVBERLOAMRARRRQGRPPSODDVEAAEEETRALYREYRTTLKRTTGQGGGIR 60
QY SSESIPAAAEBAEBCRCGPHLNRAATKSPQTPKRSQGSVPDQGRKANKLKGTLQAG 120
DB SSESIPAAAEBAEBCRCGPHLNRAATKSPQTPKRSQGSVPDQGRKANKLKGTLQAG 120
QY 61 SSESIPAAAEBAEBCRCGPHLNRAATKSPQTPKRSQGSVPDQGRKANKLKGTLQAG 120
DB 61 SSESIPAAAEBAEBCRCGPHLNRAATKSPQTPKRSQGSVPDQGRKANKLKGTLQAG 120
QY 121 PALGRPMPPLGRASSKASTPKPGTGPVPSFAEKVSDPEPPQLEPQPRGRLOHLQASIS 180
DB 121 PALGRPMPPLGRASSKASTPKPGTGPVPSFAEKVSDPEPPQLEPQPRGRLOHLQASIS 180
QY 181 QRLGSLDQGMLORCHESEVPDFLGAPKACRPDLGSEBSQULLTGESAVLGPAGSGQPEAS 240
DB 181 QRLGSLDQGMLORCHESEVPDFLGAPKACRPDLGSEBSQULLTGESAVLGPAGSGQPEAS 240
QY 241 AFOEVSIVGSPQSSSGGEEKRANEBPWPESPAVOQSSSOAGPPSEGAVAVEDPPG 300
DB 241 AFOEVSIVGSPQSSSGGEEKRANEBPWPESPAVOQSSSOAGPPSEGAVAVEDPPG 300
QY 301 EPVQAPQPPQSSSPNPRYHGLSPSSQARAGAEGTAPLHFPRLARDHGRNYVRLNKKQ 360
DB 301 EPVQAPQPPQSSSPNPRYHGLSPSSQARAGAEGTAPLHFPRLARDHGRNYVRLNKKQ 360
QY 361 EPVQAPQPPQSSSPNPRYHGLSPSSQARAGAEGTAPLHFPRLARDHGRNYVRLNKKQ 360
DB 361 EPVQAPQPPQSSSPNPRYHGLSPSSQARAGAEGTAPLHFPRLARDHGRNYVRLNKKQ 360
QY 361 KHYVGRALRSRLRKQAMKQKRRKKGCFGGGATVTTKESCFINEQPDHAAACPRPA 420
DB 361 KHYVGRALRSRLRKQAMKQKRRKKGCFGGGATVTTKESCFINEQPDHAAACPRPA 420
QY 421 SEEDTDAVGPBPVPSPOPVEPVLDPVLPLYSLSGSPGQAEFPAPVAFQALBQIGHQ 480
DB 421 SEEDTDAVGPBPVPSPOPVEPVLDPVLPLYSLSGSPGQAEFPAPVAFQALBQIGHQ 480
QY 481 FRPQGERAVMRILSGISTLLVLPFGAGKSLCYQLPALYSRRSPCLTVVSPBLSLMDQ 540
DB 481 FRPQGERAVMRILSGISTLLVLPFGAGKSLCYQLPALYSRRSPCLTVVSPBLSLMDQ 540
QY 541 VSGLPPLCKAACIHSGMTRKORESVLQKIRAAQVHMLTPBALVAGAGLPPAQLPVA 600
DB 541 VSGLPPLCKAACIHSGMTRKORESVLQKIRAAQVHMLTPBALVAGAGLPPAQLPVA 600
QY 601 FACIDEAHCLSQMSHNPFCYLRVCKVLERMGVHCFLGLTATATRTASDVAOHLAAVE 660
DB 601 FACIDEAHCLSQMSHNPFCYLRVCKVLERMGVHCFLGLTATATRTASDVAOHLAAVE 660
QY 661 EPDILHGPAPVPTNLHLVSMDBRDTQALLTLLOGRFQNDLSIIYYCNREDTERIAALL 720
DB 661 EPDILHGPAPVPTNLHLVSMDBRDTQALLTLLOGRFQNDLSIIYYCNREDTERIAALL 720
QY 721 RTCLHAANVPSSGGGAPKTTAAAYHAGMSSRRRRVQAFMGGQLRVVATVAFMGGLDR 780
DB 721 RTCLHAANVPSSGGGAPKTTAAAYHAGMSSRRRRVQAFMGGQLRVVATVAFMGGLDR 780
QY 781 PDVRAVHLGLPSPESYVQAVGRAGRDQPAHCHFLQPOEEDRELRRHHAADSTPL 840
DB 781 PDVRAVHLGLPSPESYVQAVGRAGRDQPAHCHFLQPOEEDRELRRHHAADSTPL 840
QY 841 AVKRLVQRFVPACTCTCTRPSPSEOGAVGGERPVKYPQZABQSHQAAPRRVCMGH 900
DB 841 AVKRLVQRFVPACTCTCTRPSPSEOGAVGGERPVKYPQZABQSHQAAPRRVCMGH 900
QY 901 BRALPILQULTVQALDMPBEBAIEITLICYLELHPHMHMLBLATTTTHGRINCGGPAOLQALA 960
DB 901 BRALPILQULTVQALDMPBEBAIEITLICYLELHPHMHMLBLATTTTHGRINCGGPAOLQALA 960
```

```
QY 961 HRCPELAVCLAQQLPEDEGQSSSVFPMVKLVDSMGWELASVRRALCOLQWDEHPRITGV 1020
DB 961 HRCPELAVCLAQQLPEDEGQSSSVFPMVKLVDSMGWELASVRRALCOLQWDEHPRITGV 1020
QY 1021 RRGTVLVEFSELAFHLSPGDLTAEEKQICDPLYGAVQARERQALRLRRTFOAFHSV 1080
DB 1021 RRGTVLVEFSELAFHLSPGDLTAEEKQICDPLYGAVQARERQALRLRRTFOAFHSV 1080
QY 1081 APPSGPCLGEQDERSTRLDLGRVFEESGSGPGBMDAOGPEPQARLOMDEQVR 1140
DB 1081 APPSGPCLGEQDERSTRLDLGRVFEESGSGPGBMDAOGPEPQARLOMDEQVR 1140
QY 1141 CDIROLFLRPEEKFSSRAVARIFHGIGSPCYPAQVYQODRRFWRKYLHLSFHALVGLAT 1200
DB 1141 CDIROLFLRPEEKFSSRAVARIFHGIGSPCYPAQVYQODRRFWRKYLHLSFHALVGLAT 1200
QY 1201 EELLQVAR 1208
DB 1201 EELLQVAR 1208
```

RESULT 2
ABG93387
ID ABG93387 standard; protein; 1208 AA.

XX ABG93387;

XX 20-NOV-2002 (first entry)

XX Human RecQ protein-like 4 (RECQL4).

XX Human; RecQ protein-like 4; RECQL4; infection; inflammation; cytostatic;

XX tumour formation; cancer; antiinflammatory; antimicrobial;

XX antisense therapy.

XX Homo sapiens.

XX US6436706-B1.

XX 20-AUG-2002.

XX 23-FEB-2001; 2001US-00792594.

XX 23-FEB-2001; 2001US-00792594.

XX (ISIS-) ISIS PHARM INC.

XX Ward DT, Watt AT;

XX WPI; 2002-689941/74.

XX N-PSDB; ABS68851.

XX New antisense compounds targeted to nucleic acids encoding RecQ protein-
XX like 4, useful for modulating expression of the nucleic acid and treating
XX diseases associated with expression of the nucleic acid in humans.

XX Disclosure: Col 47-56; 45pp: English.

XX The invention relates to a compound targeted to specific nucleobases of
XX RecQ protein-like 4 (RECQL4) and which hybridises and inhibits the
XX expression of RECQL4. The compound is useful for inhibiting the
XX expression of RECQL4 in cells or tissues and for treating an animal,
XX particularly a human suspected of having or being prone to a disease or
XX condition associated with expression of RECQL4. The compound is useful
XX for diagnostics, therapeutics and as a research reagent, e.g.
XX prophylactically to prevent or delay infection, inflammation or tumour
XX formation. This sequence represents human RECQL4 polypeptide

xx Sequence 1208 AA;

Query Match 100.0%; Score 6424; DB 5; Length 1208;
Best Local Similarity 100.0%; Pred. No. 0;

Matches	1208	Conservative	0	Mismatches	0	Indels	0	Gaps	0
Qy	1	MERLADVRLQAMERARRRGRPPSQDDVEAABEETRALYREYRTTKRTTGAAGGGLR	60						
Db	1	MERLADVRLQAMERARRRGRPPSQDDVEAABEETRALYREYRTTKRTTGAAGGGLR	60						
Qy	61	SESELPAABEAEPERCGPHLNRAATKSPPTPKRSQGSVPDQORLKAUKLTQLOAG	120						
Db	61	SESELPAABEAEPERCGPHLNRAATKSPPTPKRSQGSVPDQORLKAUKLTQLOAG	120						
Qy	121	PALGRPMPPLGRASSKASTPKPKPGTGPVPSFAEKVSDPEPOLPEBPGRGLQHLQASLS	180						
Db	121	PALGRPMPPLGRASSKASTPKPKPGTGPVPSFAEKVSDPEPOLPEBPGRGLQHLQASLS	180						
Qy	181	QRLGSLDGMLORCHESEVPDPLGAPKACRPDLGSEBSQLLPGBSAVLGPAGSGSPRAS	240						
Db	181	QRLGSLDGMLORCHESEVPDPLGAPKACRPDLGSEBSQLLPGBSAVLGPAGSGSPRAS	240						
Qy	241	AFOEVSIVGSPQSSSGGKKRKNBEPWESPAPVQOESSQAGPSSGAGAVAEDEPPG	300						
Db	241	AFOEVSIVGSPQSSSGGKKRKNBEPWESPAPVQOESSQAGPSSGAGAVAEDEPPG	300						
Qy	301	EPVQAOPQPCSSPNPRYHGLSPSSQARAGKAGTAPLHTFPLARHDGNYVRLANKQ	360						
Db	301	EPVQAOPQPCSSPNPRYHGLSPSSQARAGKAGTAPLHTFPLARHDGNYVRLANKQ	360						
Qy	361	KHYVGRALRSRLRKQAMKQMKRKGECFGGGGATVTTKESCFINQEPDHAACCPRA	420						
Db	361	KHYVGRALRSRLRKQAMKQMKRKGECFGGGGATVTTKESCFINQEPDHAACCPRA	420						
Qy	421	SEEDDAVGPEPLVSPQVPEVPSLDPTVLPLYSLGPSCQAEPAEPAFOALBOLGHOA	480						
Db	421	SEEDDAVGPEPLVSPQVPEVPSLDPTVLPLYSLGPSCQAEPAEPAFOALBOLGHOA	480						
Qy	481	FRPQGERAVMRLISGISTLLVLPTGAGSKLCYQLPALYSRRSPCLTVLVSPLSLMDQ	540						
Db	481	FRPQGERAVMRLISGISTLLVLPTGAGSKLCYQLPALYSRRSPCLTVLVSPLSLMDQ	540						
Qy	541	VSGLPCCIAKACIHSGMTRKQRESYLOKIRAAQVHMLTPPALVAGAGLPPAALDPPVA	600						
Db	541	VSGLPCCIAKACIHSGMTRKQRESYLOKIRAAQVHMLTPPALVAGAGLPPAALDPPVA	600						
Qy	601	FACIDEAHCLSQMSNFRPCYLRVCKVLRERGVHCFGLATATARRTASDAQHLAAAB	660						
Db	601	FACIDEAHCLSQMSNFRPCYLRVCKVLRERGVHCFGLATATARRTASDAQHLAAAB	660						
Qy	661	EPDLHGPAPVPTNLHLISVMDRDTQALLTLLOGRFPQULDSIIICYCRREDTERIALLL	720						
Db	661	EPDLHGPAPVPTNLHLISVMDRDTQALLTLLOGRFPQULDSIIICYCRREDTERIALLL	720						
Qy	721	RTCLHAANVPQSGGAPRTTAAYAHGMCSSRRRRVQRAFMQGLRVVATYAFQMGIDR	780						
Db	721	RTCLHAANVPQSGGAPRTTAAYAHGMCSSRRRRVQRAFMQGLRVVATYAFQMGIDR	780						
Qy	781	PQVRAVHLGLPPSPESYVQAVGRAGRGQPAHCHLPLQPOSEDLRELRRHYNADSTPL	840						
Db	781	PQVRAVHLGLPPSPESYVQAVGRAGRGQPAHCHLPLQPOSEDLRELRRHYNADSTPL	840						
Qy	841	AVKRLVQVFPACTCTCTRPPEBQSGAVGGERPVFKYRPOEABQLSHQAPRRVCMGH	900						
Db	841	AVKRLVQVFPACTCTCTRPPEBQSGAVGGERPVFKYRPOEABQLSHQAPRRVCMGH	900						
Qy	901	ERALPIQLTVQALDMPBEALITLLCYLSLHPHMLLELATYTHRLNCPGSPALQALA	960						
Db	901	ERALPIQLTVQALDMPBEALITLLCYLSLHPHMLLELATYTHRLNCPGSPALQALA	960						
Qy	961	HRCPLAVCLAQQLDEDPQSSSVYFPMVVKLVDSMGWELASVRALCOLQMDHDEPRGV	1020						
Db	961	HRCPLAVCLAQQLDEDPQSSSVYFPMVVKLVDSMGWELASVRALCOLQMDHDEPRGV	1020						
Qy	1021	RRTGTVLVEFSELAFLHSFGDLTAEBKQICDPLGRVQAREQALALRLRTFOAFHSV	1080						
Db	1021	RRTGTVLVEFSELAFLHSFGDLTAEBKQICDPLGRVQAREQALALRLRTFOAFHSV	1080						

Qy	1081	APPSGCLLEQODDEBSRRLKDLIGRYEEBEGGEPGMEADQGPESQARLQDMEDQVR	1140
Db	1081	APPSGCLLEQODDEBSRRLKDLIGRYEEBEGGEPGMEADQGPESQARLQDMEDQVR	1140
Qy	1141	CDIRQFLSLREPEKFSRAVARIFHIGSPCYPAQVYGQDRFPKRYLHLSFHALVGLAT	1200
Db	1141	CDIRQFLSLREPEKFSRAVARIFHIGSPCYPAQVYGQDRFPKRYLHLSFHALVGLAT	1200
Qy	1201	ELLQVAR 1208	
Db	1201	ELLQVAR 1208	

RESULT 3
ADQ21562
ID ADQ21562 standard; protein; 1208 AA.

ADQ21562;

26-AUG-2004 (first entry)

Human soft tissue sarcoma-upregulated protein - SEQ ID 4382.

soft tissue sarcoma; cytosarctic; gene therapy; vaccine; screening; human.

Homo sapiens.

WO2004048938-A2.

10-JUN-2004.

26-NOV-2003; 2003WO-US038193.

26-NOV-2002; 2002US-0429739P.

(PROT-) PROTEIN DESIGN LABS INC.

Aziz N, Ginsburg WM, Zlotnick A;

WPI; 2004-441208/41.

Early detection of soft tissue sarcoma comprises determining expression of a gene in a first soft tissue sample and a normal soft tissue sample and comparing the gene expression, also useful in treating soft tissue sarcoma.

Example 2; SEQ ID NO 4382; 210bp; English.

The invention relates to a novel method for detecting soft tissue sarcoma which comprises obtaining a first soft tissue sample from an individual and a normal soft tissue sample from the same or different individual, determining the expression of a gene in both samples and comparing the expression of the gene in both soft tissue samples, where a higher level of protein expression in the first soft tissue sample indicates the presence of soft tissue sarcoma. The method of the invention has cytostatic applications and may be useful for detecting soft tissue sarcoma, possibly via gene therapy or vaccine production. The nucleic acid sequences may be useful in diagnostic and screening applications. The current sequence is that of a human soft tissue sarcoma-upregulated protein of the invention. The current sequence is not shown within the specification per se but was submitted in CD format by the inventor.

Sequence 1208 AA;

Query Match 100.0%; Score 6424; DB 8; Length 1208;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MERLADVRLQAMERARRRGRPPSQDDVEAABEETRALYREYRTTKRTTGAAGGGLR	60
Db	1	MERLADVRLQAMERARRRGRPPSQDDVEAABEETRALYREYRTTKRTTGAAGGGLR	60


```
QY 61 SSESIPAAAEAPEDRCWCPHLNRAATKSPPTPKRSQGSVPDYGORLKANLKGTLQAG 120
Db 61 SSESIPAAAEAPEDRCWCPHLNRAATKSPPTPKRSQGSVPDYGORLKANLKGTLQAG 120
QY 121 PALGRRPMPPLGRASSKASTPKPGTGPVPSFAEKVSDEPPOLPEQPPRGRLOHLQASIS 180
Db 121 PALGRRPMPPLGRASSKASTPKPGTGPVPSFAEKVSDEPPOLPEQPPRGRLOHLQASIS 180
QY 181 QRLGSLDGMLORCCHSEVPDFLGAPKACRPDLGSEESQLLIGESAVLGPAGSOGPEAS 240
Db 181 QRLGSLDGMLORCCHSEVPDFLGAPKACRPDLGSEESQLLIGESAVLGPAGSOGPEAS 240
QY 241 AFQEVSIIVGSPQPSSSGGEKRMNEBPWESPAPQVQOESSQAGPPSEGA GAVAVEEDPPG 300
Db 241 AFQEVSIIVGSPQPSSSGGEKRMNEBPWESPAPQVQOESSQAGPPSEGA GAVAVEEDPPG 300
QY 301 EPVQAPQPPQCSSPSPNPRYHGLSPSSQARAGKAEETAPLHTFPRLARHGRNYVRLNMKQ 360
Db 301 EPVQAPQPPQCSSPSPNPRYHGLSPSSQARAGKAEETAPLHTFPRLARHGRNYVRLNMKQ 360
QY 361 KHYVGRALRSRLRKQAMKQKMRKKGCFCGGGATVTTKESCFINEQFDHMAAQCPRPA 420
Db 361 KHYVGRALRSRLRKQAMKQKMRKKGCFCGGGATVTTKESCFINEQFDHMAAQCPRPA 420
QY 421 SEEDTDAVGPBPVPSPOPVPBPSPLDPTVLPYSLGSPSGQLAETPAEVFOALBQLGHQA 480
Db 421 SEEDTDAVGPBPVPSPOPVPBPSPLDPTVLPYSLGSPSGQLAETPAEVFOALBQLGHQA 480
QY 481 FRPQGERAVMRLISGISLTLVLPFGAGKSLCYQLPALLYSRSRPLCTLVVSLTSLMDQ 540
Db 481 FRPQGERAVMRLISGISLTLVLPFGAGKSLCYQLPALLYSRSRPLCTLVVSLTSLMDQ 540
QY 541 VSGLPPLCAACIHSGMTKQKQRESVLOKIRAAQVAVMLTPPALVAGAGLPPAOLPPIVA 600
Db 541 VSGLPPLCAACIHSGMTKQKQRESVLOKIRAAQVAVMLTPPALVAGAGLPPAOLPPIVA 600
QY 601 FACIDEAHCLSQWSHNPFCYLKVCVLRERMGVHCFLGLTATATRRASDVAQHLAAVE 660
Db 601 FACIDEAHCLSQWSHNPFCYLKVCVLRERMGVHCFLGLTATATRRASDVAQHLAAVE 660
QY 661 EPDLHGAPVPPTNLHLSVMDPTDQALLTLLOGRFQNLDSIIYYCNRRBETERTIALTL 720
Db 661 EPDLHGAPVPPTNLHLSVMDPTDQALLTLLOGRFQNLDSIIYYCNRRBETERTIALTL 720
QY 721 RTCLHAANVPSSGSGRAPKTAAVYHAGMCSRRRRVQOAFMGOQLRVVAVAFMGGLDR 780
Db 721 RTCLHAANVPSSGSGRAPKTAAVYHAGMCSRRRRVQOAFMGOQLRVVAVAFMGGLDR 780
QY 781 PDVRAVLHLGLPSPFESYVQAVGRRAGDQPAHCHLPIQPGEDLRELRHVHADSTDFL 840
Db 781 PDVRAVLHLGLPSPFESYVQAVGRRAGDQPAHCHLPIQPGEDLRELRHVHADSTDFL 840
QY 841 AVKRLVQVFPACTCTCTPRPSEORGAVGGERPVKPYRQEMLOISHQAAPEPRVCMGH 900
Db 841 AVKRLVQVFPACTCTCTPRPSEORGAVGGERPVKPYRQEMLOISHQAAPEPRVCMGH 900
QY 901 ERALPIQTLVQALMDPEBAIEFLCYLELHPHMLELATTYTHCRINCPGPAOLOALA 960
Db 901 ERALPIQTLVQALMDPEBAIEFLCYLELHPHMLELATTYTHCRINCPGPAOLOALA 960
QY 961 HRCPLPVLCLAOQLPEDPGQSSSVFPMVKLVDSMGWELASVRRLCOLQMDHEPRGTG 1020
Db 961 HRCPLPVLCLAOQLPEDPGQSSSVFPMVKLVDSMGWELASVRRLCOLQMDHEPRGTG 1020
QY 1021 RRGTVAVFESFLAHLSPGDLTAEBKQICDPLVGVQAEERQALARTTPQAFISV 1080
Db 1021 RRGTVAVFESFLAHLSPGDLTAEBKQICDPLVGVQAEERQALARTTPQAFISV 1080
QY 1081 AFPSGCPCLQOQDERSTRLKDLGRYFEEBEGQEPGMEDEQGEPPQARLQDMEDQVR 1140
Db 1081 AFPSGCPCLQOQDERSTRLKDLGRYFEEBEGQEPGMEDEQGEPPQARLQDMEDQVR 1140
QY 1141 CDIROPFLSLRPEBEKSSRAVARIIPHIGISPCYPAQVYGQDRRFMRKYLHLSFHALVGLAT 1200
```

```
Db 1141 CDIROPFLSLRPEBEKSSRAVARIIPHIGISPCYPAQVYGQDRRFMRKYLHLSFHALVGLAT 1200
QY 1201 BELLQVAR 1208
Db 1201 BELLQVAR 1208

RESULT 4
AAM95050
ID AAM95050 standard; protein, 1208 AA.
XX
XX AAM95050;
AC
XX
XX 14-MAY-1999 (first entry)
DT
XX
DE Human helicase protein.
XX
XX RecQ4 gene; helicase; Werner's syndrome; Bloom's syndrome; human.
XX
XX Homo sapiens.
XX
XX W09905284-A1.
XX
XX 04-FEB-1999.
PD
XX
XX 10-JUL-1998; 98WO-JP003114.
PF
XX
XX 25-JUL-1997; 97JP-00200387.
PR
XX
XX (AGEN-) AGENE RES INST CO LTD.
PA
XX
XX Shitamoto A, Kitao S, Furuichi Y;
PI
XX
XX WPI; 1999-142939/12.
DR
XX
XX N-PSDB; AAX21656.
PT
XX
XX New human helicase gene RecQ4 - used for investigation and diagnosis of
PR helicase-implicated diseases such as Werner's syndrome.
XX
XX
XX Claim 1; Page 35-42; 67pp; Japanese.
PS
XX
XX The present sequence represents a protein having helicase activity
CC encoded by the human gene RecQ4. The gene has significant homology to the
CC Escherichia coli helicase gene (RecQ). Host cells transformed with
CC vectors comprising the RecQ4 gene are used for the recombinant expression
CC of the protein. The gene may be used for the study and diagnosis of
CC disorders in which helicase activity is involved, such as Werner's and
CC Bloom's syndromes in which mutations in the helicase gene are implicated
XX
XX
SQ Sequence 1208 AA;

Query Match 100.0%; Score 6421; DB 2; Length 1208;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1207; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MERLADVVERLOQAMERARFRORGRSPQDDVEAAPAEETRALYREYRTLKRTTGAAGGLR 60
Db 1 MERLADVVERLOQAMERARFRORGRSPQDDVEAAPAEETRALYREYRTLKRTTGAAGGLR 60
QY 61 SSESIPAAAEAPEDRCWCPHLNRAATKSPPTPKRSQGSVPDYGORLKANLKGTLQAG 120
Db 61 SSESIPAAAEAPEDRCWCPHLNRAATKSPPTPKRSQGSVPDYGORLKANLKGTLQAG 120
QY 121 PALGRRPMPPLGRASSKASTPKPGTGPVPSFAEKVSDEPPOLPEQPPRGRLOHLQASIS 180
Db 121 PALGRRPMPPLGRASSKASTPKPGTGPVPSFAEKVSDEPPOLPEQPPRGRLOHLQASIS 180
QY 181 QRLGSLDGMLORCCHSEVPDFLGAPKACRPDLGSEESQLLIGESAVLGPAGSOGPEAS 240
Db 181 QRLGSLDGMLORCCHSEVPDFLGAPKACRPDLGSEESQLLIGESAVLGPAGSOGPEAS 240
QY 241 AFQEVSIIVGSPQPSSSGGEKRMNEBPWESPAPQVQOESSQAGPPSEGA GAVAVEEDPPG 300
```

Db 241 APOEVSIRGSPQSSSGGKRMNEPESPAQVOQESSQAGPSEBGAGAVAEEDPPG 300
 Qy 301 EPNVQAPPOPCSSPSNRYHGLSPSSQADAGAEPTAPLHPRLARHNRGNVRLNMQ 360
 Db 301 EPNVQAPPOPCSSPSNRYHGLSPSSQADAGAEPTAPLHPRLARHNRGNVRLNMQ 360
 Qy 361 KHYVGRALRSRLRQAMKQKRRKKGEGCGGAGAVTTTKESEFNLNEQPDHMAAOCPRA 420
 Db 361 KHYVGRALRSRLRQAMKQKRRKKGEGCGGAGAVTTTKESEFNLNEQPDHMAAOCPRA 420
 Qy 421 SEEDTDAVGEPLVPSPPQVPEVPSIDPTVLPYLSIGSPGQLAETPAEFOALEQGHQ 480
 Db 421 SEEDTDAVGEPLVPSPPQVPEVPSIDPTVLPYLSIGSPGQLAETPAEFOALEQGHQ 480
 Qy 481 FRPGEERAVMRLISGISTLLVPTGAGKSLCTQLPALLYSRSPCLTLVSPSLISMDQ 540
 Db 481 FRPGEERAVMRLISGISTLLVPTGAGKSLCTQLPALLYSRSPCLTLVSPSLISMDQ 540
 Qy 541 VSGLPPLCAACIHSQMTKORSESVLQKTRAAQVHMLTPREALVAGGLPPAQLPPVA 600
 Db 541 VSGLPPLCAACIHSQMTKORSESVLQKTRAAQVHMLTPREALVAGGLPPAQLPPVA 600
 Qy 601 FACIDEANCLSQMSHNFRCYLKVCYLREMGVHCFGLTATATRTASDVAAQHLAVAE 660
 Db 601 FACIDEANCLSQMSHNFRCYLKVCYLREMGVHCFGLTATATRTASDVAAQHLAVAE 660
 Qy 661 EBDLHGPAPVPTNLHLSVSMRDTQALLTLQKRPQNLDSIIYCNREDTERIALLL 720
 Db 661 EBDLHGPAPVPTNLHLSVSMRDTQALLTLQKRPQNLDSIIYCNREDTERIALLL 720
 Qy 721 RFLCLAAMVPSGGGAPKTTAAVYHAGMSRRRRQVAFMGOQLRVVYATVAFMGGLR 780
 Db 721 RFLCLAAMVPSGGGAPKTTAAVYHAGMSRRRRQVAFMGOQLRVVYATVAFMGGLR 780
 Qy 781 PDVRAVHLGLPPSESVYQAVGRAGRDQPAHCHFLQPOSEDLRELRRHYNADSTDL 840
 Db 781 PDVRAVHLGLPPSESVYQAVGRAGRDQPAHCHFLQPOSEDLRELRRHYNADSTDL 840
 Qy 841 AVKRLVQVRFPACTCTTRPSEBQGAVGGERPVKTPQEAOLSHQAAPGRRVCMGH 900
 Db 841 AVKRLVQVRFPACTCTTRPSEBQGAVGGERPVKTPQEAOLSHQAAPGRRVCMGH 900
 Qy 901 BEALPIQLVQALDMPBEALFTLLCYELHPRHMLLATTYTHCLNCPGPAQLOALA 960
 Db 901 BEALPIQLVQALDMPBEALFTLLCYELHPRHMLLATTYTHCLNCPGPAQLOALA 960
 Qy 961 HRCPLAVCLAQQLPEDPGQSSSVFDMVKLVDSMGWELASVRALCOLQMDHEPRTGV 1020
 Db 961 HRCPLAVCLAQQLPEDPGQSSSVFDMVKLVDSMGWELASVRALCOLQMDHEPRTGV 1020
 Qy 1021 RRGTVLVEFSEFLAHLNSPGLTAEKQOICDPLVGRVQAEERQALARTFOAFHSV 1080
 Db 1021 RRGTVLVEFSEFLAHLNSPGLTAEKQOICDPLVGRVQAEERQALARTFOAFHSV 1080
 Qy 1081 AFPSGSPCLQOODEBRSRLKDLGRYREBEGQERGMEDQGPBGARLQMDWQVR 1140
 Db 1081 AFPSGSPCLQOODEBRSRLKDLGRYREBEGQERGMEDQGPBGARLQMDWQVR 1140
 Qy 1141 CDIROPFLSLRPEKSSRAVARIFHIGISPCYPAQVYQODRRFMRKYLLSLFHALVGLAT 1200
 Db 1141 CDIROPFLSLRPEKSSRAVARIFHIGISPCYPAQVYQODRRFMRKYLLSLFHALVGLAT 1200
 Qy 1201 EELLQVAR 1208
 Db 1201 EELLQVAR 1208

XX 03-JUN-2004 (first entry)
 DT Human pharmaceutically useful protein SegID 349.
 DE human; cancer; haematopoiesis; thrombosis; anaemia;
 XX cardiovascular disorder; ischaemic heart disease;
 XX acute myocardial infarction; respiratory disease; asthma; pneumonia;
 XX cystic fibrosis; chronic renal failure; glomerulopathy;
 XX gastrointestinal disorder; peptic ulcer; cirrhosis; immune disorder;
 XX HIV infection; systemic lupus erythematosus; endocrine system;
 XX diabetes mellitus; epilepsy; Alzheimer's disease;
 XX amyotrophic lateral sclerosis; skin disorder; psoriasis; bacterial;
 XX fungal; parasitic; viral infection; cytostatic; anticoagulant;
 XX thrombolytic; antianemic; cardiac; vasotropic; antiscclerotic;
 XX antiinflammatory; nephrotropic; antiulcer; hepatotropic;
 XX immunosuppressive; antiallergic; dermatological; antineumatic;
 XX antiarthritic; anticonvulsant; neuroprotective; nootropic;
 XX antiparasitic; antibacterial; fungicide; antiparasitic; virucidal;
 XX gene therapy; vaccine.
 OS Homo sapiens.
 XX WO2004020595-A2.
 PN 11-MAR-2004.
 PD 28-AUG-2003; 2003WO-US027107.
 PF 29-AUG-2002; 2002US-0406576P.
 XX 29-AUG-2002; 2002US-0406611P.
 XX 29-AUG-2002; 2002US-0406612P.
 XX 29-AUG-2002; 2002US-0406613P.
 XX 29-AUG-2002; 2002US-0406614P.
 XX 29-AUG-2002; 2002US-0406615P.
 XX 29-AUG-2002; 2002US-0406616P.
 XX 29-AUG-2002; 2002US-0406617P.
 XX 29-AUG-2002; 2002US-0406618P.
 XX 29-AUG-2002; 2002US-0406619P.
 XX 29-AUG-2002; 2002US-0406620P.
 XX 29-AUG-2002; 2002US-0406621P.
 XX 29-AUG-2002; 2002US-0406622P.
 XX 29-AUG-2002; 2002US-0406623P.
 XX 29-AUG-2002; 2002US-0406624P.
 XX 29-AUG-2002; 2002US-0406625P.
 XX 29-AUG-2002; 2002US-0406626P.
 XX 29-AUG-2002; 2002US-0406627P.
 XX 29-AUG-2002; 2002US-0406628P.
 XX 29-AUG-2002; 2002US-0406629P.
 XX 29-AUG-2002; 2002US-0406630P.
 XX 29-AUG-2002; 2002US-0406631P.
 XX 29-AUG-2002; 2002US-0406632P.
 XX 29-AUG-2002; 2002US-0406633P.
 XX 29-AUG-2002; 2002US-0406634P.
 XX 29-AUG-2002; 2002US-0406635P.
 XX 29-AUG-2002; 2002US-0406636P.
 XX 29-AUG-2002; 2002US-0406637P.
 XX 29-AUG-2002; 2002US-0406638P.
 XX 29-AUG-2002; 2002US-0406639P.
 XX 29-AUG-2002; 2002US-0406640P.
 XX 29-AUG-2002; 2002US-0406641P.
 XX 29-AUG-2002; 2002US-0406642P.
 XX 29-AUG-2002; 2002US-0406643P.
 XX 29-AUG-2002; 2002US-0406644P.
 XX 29-AUG-2002; 2002US-0406645P.
 XX 29-AUG-2002; 2002US-0406646P.
 XX 29-AUG-2002; 2002US-0406647P.
 XX 29-AUG-2002; 2002US-0406648P.
 XX 29-AUG-2002; 2002US-0406649P.
 XX 29-AUG-2002; 2002US-0406650P.
 XX 29-AUG-2002; 2002US-0406651P.
 XX 29-AUG-2002; 2002US-0406652P.
 XX 29-AUG-2002; 2002US-0406653P.
 XX 29-AUG-2002; 2002US-0406654P.
 XX 29-AUG-2002; 2002US-0406655P.
 XX 29-AUG-2002; 2002US-0406656P.
 XX 29-AUG-2002; 2002US-0406657P.
 XX 29-AUG-2002; 2002US-0406658P.
 XX 29-AUG-2002; 2002US-0406659P.
 XX 29-AUG-2002; 2002US-0406660P.
 XX 29-AUG-2002; 2002US-0406661P.
 XX 29-AUG-2002; 2002US-0406662P.
 XX 29-AUG-2002; 2002US-0406663P.
 XX 29-AUG-2002; 2002US-0406664P.
 XX 29-AUG-2002; 2002US-0406665P.
 XX 29-AUG-2002; 2002US-0406666P.
 XX 29-AUG-2002; 2002US-0406667P.
 XX 29-AUG-2002; 2002US-0406668P.
 XX 29-AUG-2002; 2002US-0406669P.
 XX 29-AUG-2002; 2002US-0406670P.
 XX 29-AUG-2002; 2002US-0406671P.
 XX 29-AUG-2002; 2002US-0406672P.
 XX 29-AUG-2002; 2002US-0406673P.
 XX 29-AUG-2002; 2002US-0406674P.
 XX 29-AUG-2002; 2002US-0406675P.
 XX 29-AUG-2002; 2002US-0406676P.
 XX 29-AUG-2002; 2002US-0406677P.
 XX 29-AUG-2002; 2002US-0406678P.
 XX 29-AUG-2002; 2002US-0406679P.
 XX 29-AUG-2002; 2002US-0406680P.
 XX 29-AUG-2002; 2002US-0406681P.
 XX 29-AUG-2002; 2002US-0406682P.
 XX 29-AUG-2002; 2002US-0406683P.
 XX 29-AUG-2002; 2002US-0406684P.
 XX 29-AUG-2002; 2002US-0406685P.
 XX 29-AUG-2002; 2002US-0406686P.
 XX 29-AUG-2002; 2002US-0406687P.
 XX 29-AUG-2002; 2002US-0406688P.
 XX 29-AUG-2002; 2002US-0406689P.
 XX 29-AUG-2002; 2002US-0406690P.
 XX 29-AUG-2002; 2002US-0406691P.
 XX 29-AUG-2002; 2002US-0406692P.
 XX 29-AUG-2002; 2002US-0406693P.
 XX 29-AUG-2002; 2002US-0406694P.
 XX 29-AUG-2002; 2002US-0406695P.
 XX 29-AUG-2002; 2002US-0406696P.
 XX 29-AUG-2002; 2002US-0406697P.
 XX 29-AUG-2002; 2002US-0406698P.
 XX 29-AUG-2002; 2002US-0406699P.
 XX 29-AUG-2002; 2002US-0406700P.
 XX 29-AUG-2002; 2002US-0406701P.
 XX 29-AUG-2002; 2002US-0406702P.
 XX 29-AUG-2002; 2002US-0406703P.
 XX 29-AUG-2002; 2002US-0406704P.
 XX 29-AUG-2002; 2002US-0406705P.
 XX 29-AUG-2002; 2002US-0406706P.
 XX 29-AUG-2002; 2002US-0406707P.
 XX 29-AUG-2002; 2002US-0406708P.
 XX 29-AUG-2002; 2002US-0406709P.
 XX 29-AUG-2002; 2002US-0406710P.
 XX 29-AUG-2002; 2002US-0406711P.
 XX 29-AUG-2002; 2002US-0406712P.
 XX 29-AUG-2002; 2002US-0406713P.
 XX 29-AUG-2002; 2002US-0406714P.
 XX 29-AUG-2002; 2002US-0406715P.
 XX 29-AUG-2002; 2002US-0406716P.
 XX 29-AUG-2002; 2002US-0406717P.
 XX 29-AUG-2002; 2002US-0406718P.
 XX 29-AUG-2002; 2002US-0406719P.
 XX 29-AUG-2002; 2002US-0406720P.
 XX 29-AUG-2002; 2002US-0406721P.
 XX 29-AUG-2002; 2002US-0406722P.
 XX 29-AUG-2002; 2002US-0406723P.
 XX 29-AUG-2002; 2002US-0406724P.
 XX 29-AUG-2002; 2002US-0406725P.
 XX 29-AUG-2002; 2002US-0406726P.
 XX 29-AUG-2002; 2002US-0406727P.
 XX 29-AUG-2002; 2002US-0406728P.
 XX 29-AUG-2002; 2002US-0406729P.
 XX 29-AUG-2002; 2002US-0406730P.
 XX 29-AUG-2002; 2002US-0406731P.
 XX 29-AUG-2002; 2002US-0406732P.
 XX 29-AUG-2002; 2002US-0406733P.
 XX 29-AUG-2002; 2002US-0406734P.
 XX 29-AUG-2002; 2002US-0406735P.
 XX 29-AUG-2002; 2002US-0406736P.
 XX 29-AUG-2002; 2002US-0406737P.
 XX 29-AUG-2002; 2002US-0406738P.
 XX 29-AUG-2002; 2002US-0406739P.
 XX 29-AUG-2002; 2002US-0406740P.
 XX 29-AUG-2002; 2002US-0406741P.
 XX 29-AUG-2002; 2002US-0406742P.
 XX 29-AUG-2002; 2002US-0406743P.
 XX 29-AUG-2002; 2002US-0406744P.
 XX 29-AUG-2002; 2002US-0406745P.
 XX 29-AUG-2002; 2002US-0406746P.
 XX 29-AUG-2002; 2002US-0406747P.
 XX 29-AUG-2002; 2002US-0406748P.
 XX 29-AUG-2002; 2002US-0406749P.
 XX 29-AUG-2002; 2002US-0406750P.
 XX 29-AUG-2002; 2002US-0406751P.
 XX 29-AUG-2002; 2002US-0406752P.
 XX 29-AUG-2002; 2002US-0406753P.
 XX 29-AUG-2002; 2002US-0406754P.
 XX 29-AUG-2002; 2002US-0406755P.
 XX 29-AUG-2002; 2002US-0406756P.
 XX 29-AUG-2002; 2002US-0406757P.
 XX 29-AUG-2002; 2002US-0406758P.
 XX 29-AUG-2002; 2002US-0406759P.
 XX 29-AUG-2002; 2002US-0406760P.
 XX 29-AUG-2002; 2002US-0406761P.
 XX 29-AUG-2002; 2002US-0406762P.
 XX 29-AUG-2002; 2002US-0406763P.
 XX 29-AUG-2002; 2002US-0406764P.
 XX 29-AUG-2002; 2002US-0406765P.
 XX 29-AUG-2002; 2002US-0406766P.
 XX 29-AUG-2002; 2002US-0406767P.
 XX 29-AUG-2002; 2002US-0406768P.
 XX 29-AUG-2002; 2002US-0406769P.
 XX 29-AUG-2002; 2002US-0406770P.
 XX 29-AUG-2002; 2002US-0406771P.
 XX 29-AUG-2002; 2002US-0406772P.
 XX 29-AUG-2002; 2002US-0406773P.
 XX 29-AUG-2002; 2002US-0406774P.
 XX 29-AUG-2002; 2002US-0406775P.
 XX 29-AUG-2002; 2002US-0406776P.
 XX 29-AUG-2002; 2002US-0406777P.
 XX 29-AUG-2002; 2002US-0406778P.
 XX 29-AUG-2002; 2002US-0406779P.
 XX 29-AUG-2002; 2002US-0406780P.
 XX 29-AUG-2002; 2002US-0406781P.
 XX 29-AUG-2002; 2002US-0406782P.
 XX 29-AUG-2002; 2002US-0406783P.
 XX 29-AUG-2002; 2002US-0406784P.
 XX 29-AUG-2002; 2002US-0406785P.
 XX 29-AUG-2002; 2002US-0406786P.
 XX 29-AUG-2002; 2002US-0406787P.
 XX 29-AUG-2002; 2002US-0406788P.
 XX 29-AUG-2002; 2002US-0406789P.
 XX 29-AUG-2002; 2002US-0406790P.
 XX 29-AUG-2002; 2002US-0406791P.
 XX 29-AUG-2002; 2002US-0406792P.
 XX 29-AUG-2002; 2002US-0406793P.
 XX 29-AUG-2002; 2002US-0406794P.
 XX 29-AUG-2002; 2002US-0406795P.
 XX 29-AUG-2002; 2002US-0406796P.
 XX 29-AUG-2002; 2002US-0406797P.
 XX 29-AUG-2002; 2002US-0406798P.
 XX 29-AUG-2002; 2002US-0406799P.
 XX 29-AUG-2002; 2002US-0406800P.
 XX 29-AUG-2002; 2002US-0406801P.
 XX 29-AUG-2002; 2002US-0406802P.
 XX 29-AUG-2002; 2002US-0406803P.
 XX 29-AUG-2002; 2002US-0406804P.
 XX 29-AUG-2002; 2002US-0406805P.
 XX 29-AUG-2002; 2002US-0406806P.
 XX 29-AUG-2002; 2002US-0406807P.
 XX 29-AUG-2002; 2002US-0406808P.
 XX 29-AUG-2002; 2002US-0406809P.
 XX 29-AUG-2002; 2002US-0406810P.
 XX 29-AUG-2002; 2002US-0406811P.
 XX 29-AUG-2002; 2002US-0406812P.
 XX 29-AUG-2002; 2002US-0406813P.
 XX 29-AUG-2002; 2002US-0406814P.
 XX 29-AUG-2002; 2002US-0406815P.
 XX 29-AUG-2002; 2002US-0406816P.
 XX 29-AUG-2002; 2002US-0406817P.
 XX 29-AUG-2002; 2002US-0406818P.
 XX 29-AUG-2002; 2002US-0406819P.
 XX 29-AUG-2002; 2002US-0406820P.
 XX 29-AUG-2002; 2002US-0406821P.
 XX 29-AUG-2002; 2002US-0406822P.
 XX 29-AUG-2002; 2002US-0406823P.
 XX 29-AUG-2002; 2002US-0406824P.
 XX 29-AUG-2002; 2002US-0406825P.
 XX 29-AUG-2002; 2002US-0406826P.
 XX 29-AUG-2002; 2002US-0406827P.
 XX 29-AUG-2002; 2002US-0406828P.
 XX 29-AUG-2002; 2002US-0406829P.
 XX 29-AUG-2002; 2002US-0406830P.
 XX 29-AUG-2002; 2002US-0406831P.
 XX 29-AUG-2002; 2002US-0406832P.
 XX 29-AUG-2002; 2002US-0406833P.
 XX 29-AUG-2002; 2002US-0406834P.
 XX 29-AUG-2002; 2002US-0406835P.
 XX 29-AUG-2002; 2002US-0406836P.
 XX 29-AUG-2002; 2002US-0406837P.
 XX 29-AUG-2002; 2002US-0406838P.
 XX 29-AUG-2002; 2002US-0406839P.
 XX 29-AUG-2002; 2002US-0406840P.
 XX 29-AUG-2002; 2002US-0406841P.
 XX 29-AUG-2002; 2002US-0406842P.
 XX 29-AUG-2002; 2002US-0406843P.
 XX 29-AUG-2002; 2002US-0406844P.
 XX 29-AUG-2002; 2002US-0406845P.
 XX 29-AUG-2002; 2002US-0406846P.
 XX 29-AUG-2002; 2002US-0406847P.
 XX 29-AUG-2002; 2002US-0406848P.
 XX 29-AUG-2002; 2002US-0406849P.
 XX 29-AUG-2002; 2002US-0406850P.
 XX 29-AUG-2002; 2002US-0406851P.
 XX 29-AUG-2002; 2002US-0406852P.
 XX 29-AUG-2002; 2002US-0406853P.
 XX 29-AUG-2002; 2002US-0406854P.
 XX 29-AUG-2002; 2002US-0406855P.
 XX 29-AUG-2002; 2002US-0406856P.
 XX 29-AUG-2002; 2002US-0406857P.
 XX 29-AUG-2002; 2002US-0406858P.
 XX 29-AUG-2002; 2002US-0406859P.
 XX 29-AUG-2002; 2002US-0406860P.
 XX 29-AUG-2002; 2002US-0406861P.
 XX 29-AUG-2002; 2002US-0406862P.
 XX 29-AUG-2002; 2002US-0406863P.
 XX 29-AUG-2002; 2002US-0406864P.
 XX 29-AUG-2002; 2002US-0406865P.
 XX 29-AUG-2002; 2002US-0406866P.
 XX 29-AUG-2002; 2002US-0406867P.
 XX 29-AUG-2002; 2002US-0406868P.
 XX 29-AUG-2002; 2002US-0406869P.
 XX 29-AUG-2002; 2002US-0406870P.
 XX 29-AUG-2002; 2002US-0406871P.
 XX 29-AUG-2002; 2002US-0406872P.
 XX 29-AUG-2002; 2002US-0406873P.
 XX 29-AUG-2002; 2002US-0406874P.
 XX 29-AUG-2002; 2002US-0406875P.
 XX 29-AUG-2002; 2002US-0406876P.
 XX 29-AUG-2002; 2002US-0406877P.
 XX 29-AUG-2002; 2002US-0406878P.
 XX 29-AUG-2002; 2002US-0406879P.
 XX 29-AUG-2002; 2002US-0406880P.
 XX 29-AUG-2002; 2002US-0406881P.
 XX 29-AUG-2002; 2002US-0406882P.
 XX 29-AUG-2002; 2002US-0406883P.
 XX 29-AUG-2002; 2002US-0406884P.
 XX 29-AUG-2002; 2002US-0406885P.
 XX 29-AUG-2002; 2002US-0406886P.
 XX 29-AUG-2002; 2002US-0406887P.
 XX 29-AUG-2002; 2002US-0406888P.
 XX 29-AUG-2002; 2002US-0406889P.
 XX 29-AUG-2002; 2002US-0406890P.
 XX 29-AUG-2002; 2002US-0406891P.
 XX 29-AUG-2002; 2002US-0406892P.
 XX 29-AUG-2002; 2002US-0406893P.
 XX 29-AUG-2002; 2002US-0406894P.
 XX 29-AUG-2002; 2002US-0406895P.
 XX 29-AUG-2002; 2002US-0406896P.
 XX 29-AUG-2002; 2002US-0406897P.
 XX 29-AUG-2002; 2002US-0406898P.
 XX 29-AUG-2002; 2002US-0406899P.
 XX 29-AUG-2002; 2002US-0406900P.
 XX 29-AUG-2002; 2002US-0406901P.
 XX 29-AUG-2002; 2002US-0406902P.
 XX 29-AUG-2002; 2002US-0406903P.
 XX 29-AUG-2002; 2002US-0406904P.
 XX 29-AUG-2002; 2002US-0406905P.
 XX 29-AUG-2002; 2002US-0406906P.
 XX 29-AUG-2002; 2002US-0406907P.
 XX 29-AUG-2002; 2002US-0406908P.
 XX 29-AUG-2002; 2002US-0406909P.
 XX 29-AUG-2002; 2002US-0406910P.
 XX 29-AUG-2002; 2002US-0406911P.
 XX 29-AUG-2002; 2002US-0406912P.
 XX 29-AUG-2002; 2002US-0406913P.
 XX 29-AUG-2002; 2002US-0406914P.
 XX 29-AUG-2002; 2002US-0406915P.
 XX 29-AUG-2002; 2002US-0406916P.
 XX 29-AUG-2002; 2002US-0406917P.
 XX 29-AUG-2002; 2002US-0406918P.
 XX 29-AUG-2002; 2002US-0406919P.
 XX 29-AUG-2002; 2002US-0406920P.
 XX 29-AUG-2002; 2002US-0406921P.
 XX 29-AUG-2002; 2002US-0406922P.
 XX 29-AUG-2002; 2002US-0406923P.
 XX 29-AUG-2002; 2002US-0406924P.
 XX 29-AUG-2002; 2002US-0406925P.
 XX 29-AUG-2002; 2002US-0406926P.
 XX 29-AUG-2002; 2002US-0406927P.
 XX 29-AUG-2002; 2002US-0406928P.
 XX 29-AUG-2002; 2002US-0406929P.
 XX 29-AUG-2002; 2002US-0406930P.
 XX 29-AUG-2002; 2002US-0406931P.
 XX 29-AUG-

CC polymyocytoides, vectors and host cells useful for diagnosing, preventing
 CC and treating proliferative disorders, e.g. cancer, disorders of
 CC haematopoiesis such as thrombosis and anaemia, cardiovascular disorders,
 CC e.g. ischaemic heart disease and acute myocardial infarction, respiratory
 CC diseases, e.g. asthma, pneumonia or cystic fibrosis, disorders of the
 CC kidney and urinary tract, e.g. chronic renal failure and glomerulopathy,
 CC gastroenteric disorders, e.g. peptic ulcer or cirrhosis, immune
 CC disorders, e.g. HIV infection and systemic lupus erythematosus, disorders
 CC of the endocrine system, e.g. diabetes mellitus, central nervous system
 CC disorders, e.g. epilepsy, Alzheimer's disease or anyotropic lateral
 CC sclerosis, skin disorders, e.g. psoriasis, as well as bacterial, fungal,
 CC parasitic and viral diseases. Accordingly, they exhibit many various
 CC activities including cytostatic, anticoagulant, thrombolytic,
 CC antianemic, cardiac, vasotropic, antianasthetic, antiinflammatory,
 CC nephrotropic, antitumor, hepatotropic, immunosuppressive, antiallergic,
 CC dermatological, antineumatic, antiarthritic, antidiabetic,
 CC anticonvulsant, neuroprotective, nootropic, antipsoriatic, antibacterial,
 CC fungicide, antiparasitic and virucidal, such that these polymyocytoides
 CC can be used for gene therapy purposes and the development of appropriate
 CC vaccines. This polypeptide is a human protein of the invention.

XX Sequence 1306 AA:

Query Match 94.7%; Score 6085.5; DB 8; Length 1306;
 Best Local Similarity 92.6%; Pred. No. 0;
 Matches 1167; Conservative 2; Mismatches 26; Indels 65; Gaps 6;

QY 1 MERLRDVRRLQWENAFRRQRRPSODVNAPEETRLREYTKTKTQAGGSR 60
 DB 60 MERLRDVRRLQWENAFRRQRRPSODVNAPEETRLREYTKTKTQAGGSR 119
 QY 61 SSESPLAAEBAEPEPCMGPHLNARATKSPPTPGSRQSVDPYQRLKANLKTLOAG 120
 DB 120 SSESPLAAEBAEPEPCMGPHLNARATKSPPTPGSRQSVDPYQRLKANLKTLOAG 179
 QY 121 PALGRRPWPLGRASSASTPKPGTGPVPSFAEKVSDPEPQLPEPGRRLQHLQASIS 180
 DB 180 PALGRRPWPLGRASSASTPKPGTGPVPSFAEKVSDPEPQLPEPGRRLQHLQASIS 239
 QY 181 QRLGSLDPLQCHSEVDFDLPAPKACRPDLGSESSQLIPESAVLGPAGSOGPEAS 240
 DB 240 QRLGSLDPLQCHSEVDFDLPAPKACRPDLGSESSQLIPESAVLGPAGSOGPEAS 299
 QY 241 AFOEVSIRVSGSPSSSGGKKRWNEPEWSPPAVOQESSQAPPEEGAGAVAEEDPPG 300
 DB 300 AFOEVSIRVSGSPSSSGGKKRWNEPEWSPPAVOQESSQAPPEEGAGAVAEEDPPG 359
 QY 301 EPVQAQPPQPCSSPNRYHGLSPSSQARAGKAGTAPLHIFPRLARHGRNVRLNMKQ 360
 DB 360 EPVQAQPPQPCSSPNRYHGLSPSSQARAGKAGTAPLHIFPRLARHGRNVRLNMKQ 419
 QY 361 KHYVGRALRSRLRKQAMKQKKRKGEGCGGAGAVTTTKESFTLNEQDPHMAAQCPRP 420
 DB 420 KHYVGRALRSRLRKQAMKQKKRKGEGCGGAGAVTTTKESFTLNEQDPHMAAQCPRP 479
 QY 421 SEEDTAVGPEPLVPSPQVPEVPSDPTVLPLYSGISGQALAEPTAEVFOALEQGHQA 480
 DB 480 SEEDTAVGPEPLVPSPQVPEVPSDPTVLPLYSGISGQALAEPTAEVFOALEQGHQA 539
 QY 481 FRPGQRAVWRILISGISTLLVPTGAGKSLCYQLPALVSRSPCLTLVSPILSLMDQ 540
 DB 540 FRPGQRAVWRILISGISTLLVPTGAGKSLCYQLPALVSRSPCLTLVSPILSLMDQ 599
 QY 541 VSGLPCLCAACHSMTKQKQRESVQKTRAAOVHTLMTLTPALVAGGILPPAAQLPPVA 600
 DB 600 VSGLPCLCAACHSMTKQKQRESVQKTRAAOVHTLMTLTPALVAGGILPPAAQLPPVA 659
 QY 601 PACIDBAHCLSQSHNFRPCYLVCKVLERMGVHFGTLTAATARTASDVQCHLAAYE 660
 DB 660 PACIDBAHCLSQSHNFRPCYLVCKVLERMGVHFGTLTAATARTASDVQCHLAAYE 719
 QY 661 EPDLHGPAPVPTNLHLSVSMRDTQALLTLQGRFQNLDSIIYCNREDTERIAALL 720

DB 720 EPDLHGPAPVPTNLHLSVSMRDTQALLTLQGRFQNLDSIIYCNREDTERIAALL 779
 QY 721 RFLCHAAWPGSGGAPKTTAAVHAGMSRRRRVORAFOGOLR----- 766
 DB 780 RFLCHAAWPGSGGAPKTTAAVHAGMSRRRRVORAFOGOLRGGGGLDAGCP 839
 QY 767 -VVAVTVAFGMLDPPVAVLHLGLPPSFESYVAVAGRAGDGPANLHFLQP----- 820
 DB 840 AHCACAAASGAA---FKURE-LRAGRPF-----GRAPRLAPPPTLVLEPPVVAH 886
 QY 821 -----QGEDELRELRHHVADSTPLAVKRLVGVFPACTCTCTPPSEOGANGGERP 873
 DB 887 TLMKVALQGEDELRELRHHVADSTPLAVKRLVGVFPACTCTCTPPSEOGANGGERP 946
 QY 874 VKYRPOAEQOLSHOAAPEPRRVCMGHERALPIQLTVQALDMPDEE----- 918
 DB 947 VKYRPOAEQOLSHOAAPEPRRVCMGHERALPIQLTVQALDMPDEE----- 1006
 QY 919 -----ALETLLCYLELHPHNMLELATYTHCRINCPGPAQLQALAHRCPLAV 968
 DB 1007 PRPLSPALPRALETLLCYLELHPHNMLELATYTHCRINCPGPAQLQALAHRCPLAV 1066
 QY 969 CLAQQLPEDPGGSSSVFEDMYKLYDSMGWELASVRALCOLOMHEPRTYRGTGVAV 1028
 DB 1067 CLAQQLPEDPGGSSSVFEDMYKLYDSMGWELASVRALCOLOMHEPRTYRGTGVAV 1126
 QY 1029 EFSSELAFHLRSPGDLTAEEKDQICDPLVGRVQAREEQALARLRTPFOAHSAVAFPSGCPG 1088
 DB 1127 EFSSELAFHLRSPGDLTAEEKDQICDPLVGRVQAREEQALARLRTPFOAHSAVAFPSGCPG 1186
 QY 1089 LEQDEBERSTRKDLGLGRYFEEBEGQEPGMDAOGPEPQARLQDMEQVACDTRQFUS 1148
 DB 1187 LEQDEBERSTRKDLGLGRYFEEBEGQEPGMDAOGPEPQARLQDMEQVACDTRQFUS 1246
 QY 1149 LRPEKFSRAVARIFHGIGSPCYPAQVYGQDRFRKRYLHLSFALVGLATEBELLQVAR 1208
 DB 1247 LRPEKFSRAVARIFHGIGSPCYPAQVYGQDRFRKRYLHLSFALVGLATEBELLQVAR 1306

RESULT 6
 ADD01202
 ID ADD01202 standard; protein; 1144 AA.
 XX
 AC ADD01202;
 XX
 DT 01-JAN-2004 (first entry)
 XX
 XX Human nucleic acid-associated protein NAAP-40 SEQ ID NO:40.
 DE
 XX human; nucleic acid-associated protein; NAAP; cytosolic;
 KW antiarteriosclerotic; anti-HIV; antiallergic; cerebroprotective;
 KW antiparkinsonian; anticonvulsant; nootropic; neuroprotective;
 KW antiinflammatory; ophtalmological; thyromimetic; antiarthritic;
 KW hepatotropic; antibacterial; virucide; protozoacide; antiparasitic;
 KW fungicide; gene therapy; cell proliferative disease; cancer;
 KW atherosclerosis; hepatitis; neurological disorder; Parkinson's disease;
 KW Alzheimer's disease; stroke; epilepsy; developmental disorder;
 KW renal tubular acidosis; anaemia; glaucoma; hypochyroidism;
 KW autoimmune disorder; inflammatory disorder; AIDS; allergy;
 KW atopic dermatitis; arthritis; infection.
 KW
 XX Homo sapiens.
 OS
 XX
 PN MO2003054219-A2.
 XX
 PD 03-JUL-2003.
 XX
 PF 18-DEC-2002; 2002WO-US041115.
 XX
 PR 19-DEC-2001; 2001US-0343004P.
 XX
 PR 11-JAN-2002; 2002US-0347633P.
 XX
 PR 25-JAN-2002; 2002US-0351749P.
 XX
 PR 22-FEB-2002; 2002US-0359498P.

XX (INCYTE GENOMICS INC.

PA Azimzai Y, Baughn MR, Becha SD, Borowsky ML, Burford N;
 PI Elliott VS, Emerling BM, Forsythe ID, Goryard AE, Griffin JA;
 PI Kable AB, Khare R, Lai PG, Lee SA, Lee SY, Li JX, Marquis JP;
 PI Ramumur J, Richardson TW, Sprague WM, Swarnakar A, Tang YT;
 PI Chawla NK, Warren BA, Yue H;

DR WPI, 2003-559157/52.
 DR N-PSDB; ADD01259.

PT New human nucleic acid-associated proteins (NAAP), useful for diagnosing,
 PT treating and preventing diseases or conditions associated with the
 PT aberrant NAAP expression, e.g. cancer, hepatitis, AIDS, atherosclerosis,
 PT infections.

PS Claim 1; SEQ ID NO 40; 405bp; English.

XX The present invention describes human nucleic acid-associated proteins
 CC designated NAAP-1 to NAAP-57. The human NAAPs have cytostatic,
 CC antiarteriosclerotic, anti-HIV, antiallergic, cerebroprotective,
 CC antiParkinsonian, anticonvulsant, nootropic, neuroprotective,
 CC antiinflammatory, ophthalmological, rheumatologic, antirheumatic,
 CC hepatotropic, antibacterial, virucide, protozoacide, antiparasitic and
 CC fungicide activities, and can be used in gene therapy. The NAAP protein
 CC and polynucleotide sequences can be used in diagnosing, treating and
 CC preventing diseases or conditions associated with the decreased
 CC expression or overexpression of NAAP, such as cell proliferative diseases
 CC (e.g. cancer, atherosclerosis, hepatitis), neurological disorders
 CC (Parkinson's disease, Alzheimer's disease, stroke, epilepsy),
 CC developmental disorders (renal tubular acidosis, anaemia, glaucoma,
 CC hypochylolism), autoimmune/inflammatory disorders (AIDS, allergies,
 CC atopic dermatitis, arthritis) and infections (e.g. bacterial, viral,
 CC parasitic, protozoal, fungal). The present sequence represents human NAAP
 CC -40, from the present invention.

XX Sequence 1144 AA;

Query Match 94.2%; Score 6050; DB 7; Length 1144;
 Best Local Similarity 94.7%; Pred. No. 0;
 Matches 1144; Conservative 0; Mismatches 0; Indels 64; Gaps 1;

QY 1 MERLDVPERLQAWERAFRRGRRPSODVAAPEBTALYREYTLKRTTGGAGGGR 60
 DB 1 MERLDVPERLQAWERAFRRGRRPSODVAAPEBTALYREYTLKRTTGGAGGGR 60
 QY 61 SSESIPAAAEAPBPRCWPFLNRAATKSPPTPGSRQGSVPDYQRLKANKLKTLOAG 120
 DB 61 SSESIPAAAEAPBPRCWPFLNRAATKSPPTPGSRQGSVPDYQRLKANKLKTLOAG 120
 QY 61 SSESIPAAAEAPBPRCWPFLNRAATKSPPTPGSRQGSVPDYQRLKANKLKTLOAG 120
 DB 121 PALGRPMPFLGRASSKASTPKPGTGPVPSFAEKVSDDEPPQLPEPQPRGRLOHLQASIS 180
 DB 121 PALGRPMPFLGRASSKASTPKPGTGPVPSFAEKVSDDEPPQLPEPQPRGRLOHLQASIS 180
 QY 181 QRLGSLDPCWMLORCHSEVPDPLGAPACRPDLGSESSQLITIGESAVLGPAGSQSPAS 240
 DB 181 QRLGSLDPCWMLORCHSEVPDPLGAPACRPDLGSESSQLITIGESAVLGPAGSQSPAS 240
 QY 241 AFOEVSIRVGSPOPSSSGEKKRMNEEPWESPAVOQESSQAQPSBAGAVAVEEDPPG 300
 DB 241 AFOEVSIRVGSPOPSSSGEKKRMNEEPWESPAVOQESSQAQPSBAGAVAVEEDPPG 300
 QY 301 BEVQAQPPQPCSSPNRYHGLSPSSQARAGAEGTAPLHIFPRLARHDGRNVYRLNMQ 360
 DB 301 BEVQAQPPQPCSSPNRYHGLSPSSQARAGAEGTAPLHIFPRLARHDGRNVYRLNMQ 360
 QY 361 KHYVGRALRSRLTKQAKQKRRKKGFGGGAATVTTKESCFINQEDHMAAOCPRPA 420
 DB 361 KHYVGRALRSRLTKQAKQKRRKKGFGGGAATVTTKESCFINQEDHMAAOCPRPA 420
 QY 421 SEEDTDVAGPEPLVPSPPQVPEVPSLDPTVLPLYSLSGSQLAETPAEYFOALEQLGHQA 480
 DB 421 SEEDTDVAGPEPLVPSPPQVPEVPSLDPTVLPLYSLSGSQLAETPAEYFOALEQLGHQA 480

DB 421 SEEDTDVAGPEPLVPSPPQVPEVPSLDPTVLPLYSLSGSQLAETPAEYFOALEQLGHQA 480
 QY 481 FRPGGERAVMRILSGISTLLVLPFGAGKSLCYQLPALYSRSPCLTIVVSLSLMDQ 540
 DB 481 FRPGGERAVMRILSGISTLLVLPFGAGKSLCYQLPALYSRSPCLTIVVSLSLMDQ 540
 QY 541 VSGLPCLKAACIHSGMRKORESVLOKIRAAQVAVMLTPALVAGAGLPPAOLPVA 600
 DB 541 VSGLPCLKAACIHSGMRKORESVLOKIRAAQVAVMLTPALVAGAGLPPAOLPVA 600
 QY 601 FACIDEANCLQWSNHPFCYLAVCKVLREMGVHCFLGTATATRTASDVAOHLAVAE 660
 DB 601 FACIDEANCLQWSNHPFCYLAVCKVLREMGVHCFLGTATATRTASDVAOHLAVAE 660
 QY 661 EPDLHGPAVPFNHLSVSMDDPTQALLTLTGKRFQNLDSIIYCNREDTERIAALL 720
 DB 661 EPDLHGPAVPFNHLSVSMDDPTQALLTLTGKRFQNLDSIIYCNREDTERIAALL 720
 QY 721 RTCLHAAMVPGSGRAPKTTAEAYHAGMCSRERRVQAIFMGQLRVVAVAFMGGLDR 780
 DB 721 RTCLHAAMVPGSGRAPKTTAEAYHAGMCSRERRR----- 755
 QY 781 PDVRAVHLGLPPSESYVQAVGRAGDQPAHCHLFLQPGEDLRELRHVHADSTDEL 840
 DB 756 -----PQEDLRELRHVHADSTDEL 776
 QY 841 AVKRLVORVFPACTCTTRPSEBQAVGGERPVKYPPOEBOULSHQAPPRVCMGH 900
 DB 777 AVKRLVORVFPACTCTTRPSEBQAVGGERPVKYPPOEBOULSHQAPPRVCMGH 836
 QY 901 EBALPIQLTVQALNDPBEAIEITLLCYELHPHMELEATTYTHRLNCPGSPALQOLA 960
 DB 837 EBALPIQLTVQALNDPBEAIEITLLCYELHPHMELEATTYTHRLNCPGSPALQOLA 896
 QY 961 HRCPLAVCLAQQLPBDPQGSSEVFPDWVKLVDSMGWELASVRBALCOLQWDHEPRTGV 1020
 DB 897 HRCPLAVCLAQQLPBDPQGSSEVFPDWVKLVDSMGWELASVRBALCOLQWDHEPRTGV 956
 QY 1021 RRGTVLVFESFLAHLASPGDLTAEKQDICTFLYGRVQAEBOALRLRTFOAFHV 1080
 DB 957 RRGTVLVFESFLAHLASPGDLTAEKQDICTFLYGRVQAEBOALRLRTFOAFHV 1016
 QY 1081 APPSGCPCLEQDERSTRKDLGRYFEESEGOEPGMEHQGPPOGARLODMEDQVR 1140
 DB 1017 APPSGCPCLEQDERSTRKDLGRYFEESEGOEPGMEHQGPPOGARLODMEDQVR 1076
 QY 1141 CDIROLSLRPEKESRAVARIFFIGISPCYPAQVYGQDRRFWRKYLHLSFHALVGLAT 1200
 DB 1077 CDIROLSLRPEKESRAVARIFFIGISPCYPAQVYGQDRRFWRKYLHLSFHALVGLAT 1136
 QY 1201 BELLOVAR 1208
 DB 1137 BELLOVAR 1144

RESULT 7
 ABM64821
 ID ABM64821 standard; protein; 1142 AA.
 XX
 XX ABM64821;
 DE 18-NOV-2004 (first entry)
 XX
 DE Human diagnostic and therapeutic protein SEQ ID NO:5070.
 XX
 KW gene therapy; human diagnostic and therapeutic polynucleotide; dltip.
 XX
 OS Homo sapiens.
 XX
 PN WO2004023973-A2.
 XX
 PD 25-MAR-2004.
 XX

QY	36	KHYVGRALRSRLRQAKQKMRKKGEFFGGGATVTTKSSCFINBEPDIMAQCRRPA	420
Db	361	KHYVGRALRSRLRQAKQKMRKKGEFFGGGATVTTKSSCFINBEPDIMAQCRRPA	420
QY	421	SEEDTAVGEBPVSPQPEVPVPSIDPTVLPVLSGPSQGLAETPAEVFOALEQCHQA	480
Db	421	SEEDTAVGEBPVSPQPEVPVPSIDPTVLPVLSGPSQGLAETPAEVFOALEQCHQA	480
QY	481	FRQGEBAWVRILISGISTLLVLPPTAGAGSLCYQPLALYSRRSPCLTVVPSILSLMDQ	540
Db	481	FRQGEBAWVRILISGISTLLVLPPTAGAGSLCYQPLALYSRRSPCLTVVPSILSLMDQ	540
QY	541	VSGLPPLCAKACIHSGMTKRESESVLQKIRAAQVHVLMTPEALVGAAGLPRAAQLPPVA	600
Db	541	VSGLPPLCAKACIHSGMTKRESESVLQKIRAAQVHVLMTPEALVGAAGLPRAAQLPPVA	600
QY	601	PACIDEAHCLISQMSHNPRPCYRVCKVLEBENGHCFCGLTATATRASVAQHLVAE	660
Db	601	PACIDEAHCLISQMSHNPRPCYRVCKVLEBENGHCFCGLTATATRASVAQHLVAE	660
QY	661	EPDLHGAPVPTVTLNHSVMDRDTQALLTLTLOGRPONLDSIIYYCRREDTERIAAL	720
Db	661	EPDLHGAPVPTVTLNHSVMDRDTQALLTLTLOGRPONLDSIIYYCRREDTERIAAL	720
QY	721	RTCLHAAMVWGSGRAPKTTAAVYAHGMSCRERRR	780
Db	721	RTCLHAAMVWGSGRAPKTTAAVYAHGMSCRERRR	780
QY	781	PDVRAVHLHGLPPSPFSVQAVAGRAGDQPAHCHLFLQPOGEDRELRRHVHADSTDFL	840
Db	781	PDVRAVHLHGLPPSPFSVQAVAGRAGDQPAHCHLFLQPOGEDRELRRHVHADSTDFL	840
QY	841	AVKRLVORVFPACTCTRPSPBEGAVGGERPVEKYPPOEAQLSHQABGPRVCMGH	900
Db	841	AVKRLVORVFPACTCTRPSPBEGAVGGERPVEKYPPOEAQLSHQABGPRVCMGH	900
QY	901	ERALPIQLTVQALDMEBEALITLLCYLELHPHMLLELATTYTHORLNCPEGSPQALQALA	960
Db	901	ERALPIQLTVQALDMEBEALITLLCYLELHPHMLLELATTYTHORLNCPEGSPQALQALA	960
QY	961	HRCPPPLAVCLAQQLPBDPCQSSSVSEFDWVKLVDSMGWELASVRRALCOLQMDHEPRITGV	1020
Db	961	HRCPPPLAVCLAQQLPBDPCQSSSVSEFDWVKLVDSMGWELASVRRALCOLQMDHEPRITGV	1020
QY	1021	RRTGTVLVEFSIELAFHLRSPGDIITAEKKQICDPLVGRVQAREQALRLRRTQATHSV	1080
Db	1021	RRTGTVLVEFSIELAFHLRSPGDIITAEKKQICDPLVGRVQAREQALRLRRTQATHSV	1080
QY	1081	AFBSCGCCLQOQEBESTRLKDLGYPFEEBEGQSPGMEBDAQSPBPQALQDMEQVVR	1140
Db	1081	AFBSCGCCLQOQEBESTRLKDLGYPFEEBEGQSPGMEBDAQSPBPQALQDMEQVVR	1140
QY	1141	CDIRQPLSLRPEKFFSSRAVARIFHGIGSPCYPAQVYGQDRRFWRKYLHLSFHALVGLAT	1200
Db	1141	CDIRQPLSLRPEKFFSSRAVARIFHGIGSPCYPAQVYGQDRRFWRKYLHLSFHALVGLAT	1200
QY	1201	PELLQVAR 1208	
Db	1201	PELLQVAR 1208	
QY	1201	PELLQVAR 1142	
Db	1201	PELLQVAR 1142	
RESULT 8			
AAB20995			
ID	AAB20995 standard; protein; 756 AA.		
XX	AAB20995;		
AC	AAB20995;		
XX	11-DEC-2000 (first entry)		
DT	11-DEC-2000 (first entry)		
XX	Human RecQ4 helicase mutant, mut-2.		
DE	Human RecQ4 helicase mutant, mut-2.		
XX	RecQ4 helicase; human; Rothmund-Thomson syndrome; chromosome 8q24.3;		
XX	poikiloderma congenitale; autosomal recessive; skin disorder;		
KW	poikiloderma congenitale; autosomal recessive; skin disorder;		

dermatology; antibody; prenatal diagnosis; gene therapy; mutant; mutein.

XX OS Homo sapiens.

XX PN WO20043522-A1.

XX PD 27-JUL-2000.

XX PF 19-JAN-2000; 2000WO-JP000233.

XX PR 19-JAN-1999; 99JP-00011218.

XX PA (AGEN-) AGENE RES INST CO LTD.

XX PI Kitao S, Shimamoto A, Furuichi Y;

XX DR WPI; 2000-524241/47.

XX DR N-PSDB; AAA72365.

XX PT RecQ4 helicase gene, gene products and antibody, used in the diagnosis and treatment of Rothmund-Thomson syndrome, e.g. by gene therapy.

XX PS Example 3; Page; 115pp; Japanese.

The present sequence represents a mutant human RecQ4 helicase, mut-2. The cDNA encoding sequence contains a C to T substitution relative to the cDNA encoding the wild-type RecQ4 helicase (AAA72321), which changes a Gln codon (CAG) to a stop codon (TGA), causing premature truncation of the encoded protein. The invention relates to the genomic DNA sequence of human RecQ4 helicase (AAA72320). Mutations in this gene, located on chromosome 8q24.3, are the cause of Rothmund-Thomson syndrome (also known as poikiloderma congenitale), an autosomal recessive skin disorder principally occurring in females and often accompanied by juvenile cataracts, saddle nose, congenital bone defects, hypogonadism and disturbances in the growth of hair, nails and teeth. The invention also relates to vectors and host cells comprising the human RecQ4 helicase genomic sequence. It additionally encompasses use of the RecQ4 helicase protein as a therapeutic and anti-RecQ4 antibodies as diagnostic agents. The RecQ4 helicase gene and its products, and anti-RecQ4 helicase antibodies are useful in the diagnosis, especially prenatal diagnosis, and treatment of Rothmund-Thomson syndrome. The genomic sequence may especially be used in gene therapy for this condition. Note: The present sequence is not shown in the specification, but is derived from the wild-type human RecQ4 helicase shown on pages 83-92

XX SQ Sequence 756 AA;

Query Match 62.4%; Score 4006; DB 3; Length 756;
 Best Local Similarity 100.0%; Pred. No. 2.3e-284;
 Matches 756; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MERLDVVERLQAMERAFRRQRGRPSQDDVEAABETRALYREYTLKRTTGAAGGGR 60
 DB 1 MERLDVVERLQAMERAFRRQRGRPSQDDVEAABETRALYREYTLKRTTGAAGGGR 60
 OY 61 SSESIPAAAEERPERCWPMLNRAATKSPOTTPGSRGSPVDVQRLKATLKTLOAG 120
 DB 61 SSESIPAAAEERPERCWPMLNRAATKSPOTTPGSRGSPVDVQRLKATLKTLOAG 120
 OY 121 PALGRPMPPLGRASSKASTPPKPGTGVPSFAEKVSDPEPPQPEPPRGRLOHQAAS 180
 DB 121 PALGRPMPPLGRASSKASTPPKPGTGVPSFAEKVSDPEPPQPEPPRGRLOHQAAS 180
 OY 181 ORLSGLDPGWLORCHSEVPDPLGAPKACRPDLGSEBSQLITQESAVLGPAGSGQPEAS 240
 DB 181 ORLSGLDPGWLORCHSEVPDPLGAPKACRPDLGSEBSQLITQESAVLGPAGSGQPEAS 240
 OY 241 AFOEVSIRGSPSPSSGGEKRWNEEPESPAPVOGESSQAGPSEGGAVAVEEDPPG 300
 DB 241 AFOEVSIRGSPSPSSGGEKRWNEEPESPAPVOGESSQAGPSEGGAVAVEEDPPG 300
 OY 301 BEVQAPQPPCCSPSPNRYHGLSPSSQAPAGAEATPLHIFPRLARHDRGNVYRLNMKQ 360

DB 301 BEVQAPQPPCCSPSPNRYHGLSPSSQAPAGAEATPLHIFPRLARHDRGNVYRLNMKQ 360
 OY 361 KHYVGRALRSRLLRKQAMKQKRRKGBECFGGGATVTTKESCPLNEQPDHMAAQCPRPA 420
 DB 361 KHYVGRALRSRLLRKQAMKQKRRKGBECFGGGATVTTKESCPLNEQPDHMAAQCPRPA 420
 OY 421 SEEDTDVAGPEPLVSPPOPVEPVSIDPTVLPLYSIGSPSGQLAETPAEYFOALEQSHQA 480
 DB 421 SEEDTDVAGPEPLVSPPOPVEPVSIDPTVLPLYSIGSPSGQLAETPAEYFOALEQSHQA 480
 OY 481 FRPGERAVMRILSGISTLLVLPFGAGSLCYQLPALVSRSPCLTLVSPSLIMDDQ 540
 DB 481 FRPGERAVMRILSGISTLLVLPFGAGSLCYQLPALVSRSPCLTLVSPSLIMDDQ 540
 OY 541 VSGLPPLKAACTHSGMTRKQRESYLQTRAAQVHVMILTEPALVAGAGLPPAAQLPPPA 600
 DB 541 VSGLPPLKAACTHSGMTRKQRESYLQTRAAQVHVMILTEPALVAGAGLPPAAQLPPPA 600
 OY 601 FACIDBAHCLSQWSNFRPCYLRVCKVLRERGVHCFGLTATATRTASDVAOHLAVAE 660
 DB 601 FACIDBAHCLSQWSNFRPCYLRVCKVLRERGVHCFGLTATATRTASDVAOHLAVAE 660
 OY 661 EPDLHGPAPVPTNLHLVSVMRDTPQALLVTLQGRFQNLDSIIYCNRRREDTERIAALL 720
 DB 661 EPDLHGPAPVPTNLHLVSVMRDTPQALLVTLQGRFQNLDSIIYCNRRREDTERIAALL 720
 OY 721 RTCLHAANVPSSGGRAPKTTAAHYAGCRRRRRV 756
 DB 721 RTCLHAANVPSSGGRAPKTTAAHYAGCRRRRRV 756

RESULT 9

AAB20994
 ID AAB20994 standard; protein; 554 AA.

XX AAB20994;

XX 11-DEC-2000 (first entry)

XX Human RecQ4 helicase mutant, mut-1.

XX RecQ4 helicase; human; Rothmund-Thomson syndrome; chromosome 8q24.3;

XX poikiloderma congenitale; autosomal recessive; skin disorder;

XX dermatology; antibody; prenatal diagnosis; gene therapy; mutant; mutein.

XX OS Homo sapiens.

XX WO20043522-A1.

XX 27-JUL-2000.

XX 19-JAN-2000; 2000WO-JP000233.

XX 19-JAN-1999; 99JP-00011218.

XX (AGEN-) AGENE RES INST CO LTD.

XX Kitao S, Shimamoto A, Furuichi Y;

XX WPI; 2000-524241/47.

XX N-PSDB; AAA72364.

XX RecQ4 helicase gene, gene products and antibody, used in the diagnosis and treatment of Rothmund-Thomson syndrome, e.g. by gene therapy.

XX Example 3; Page; 115pp; Japanese.

XX The present sequence represents a mutant human RecQ4 helicase, mut-1. This cDNA encoding sequence contains a 7 bp deletion relative to the cDNA encoding the wild-type RecQ4 helicase (AAA72321), which alters the reading frame and causes premature truncation of the encoded protein. The invention relates to the genomic DNA sequence of human RecQ4 helicase (AAA72320). Mutations in this gene, located on chromosome 8q24.3, are the

CC cause of Rothmund-Thomson syndrome (also known as poikiloderma
CC congenitale), an autosomal recessive skin disorder principally occurring
CC in females and often accompanied by juvenile cataracts, saddle nose,
CC congenital bone defects, hypogonadism and disturbances in the growth of
CC hair, nails and teeth. The invention also relates to vectors and host
CC cells comprising the human RecQ4 helicase genomic sequence. It
CC additionally encompasses use of the RecQ4 helicase protein as a
CC therapeutic and anti-RecQ4 antibodies as diagnostic agents. The RecQ4
CC helicase gene and its products, and anti-RecQ4 helicase antibodies are
CC useful in the diagnosis, especially prenatal diagnosis, and treatment of
CC Rothmund-Thomson syndrome. The genomic sequence may especially be used in
CC gene therapy for this condition. Note: The present sequence is not shown
CC in the specification, but is derived from the wild-type human RecQ4
CC helicase shown on pages 83-92

XX Sequence 554 AA;

Query Match 45.5%; Score 2920; DB 3; Length 554;
Best Local Similarity 100.0%; Pred. No. 5,6e-205;
Matches 550; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MERLDVRERLQAWERAFRRGRGRRSDVAAARETALREYTLRTTQAGGGR 60
DB 1 MERLDVRERLQAWERAFRRGRGRRSDVAAARETALREYTLRTTQAGGGR 60
QY 61 SSESIPAAAEAREPERCWPMLNRAATKSPPTPGRSRGSVDYQRLKANKTLOAG 120
DB 61 SSESIPAAAEAREPERCWPMLNRAATKSPPTPGRSRGSVDYQRLKANKTLOAG 120
QY 121 PALGRPPWPLGRASSAKSTPKPGTGPVPSFAKVSDEPPQLEPPQPRGRLOHLOASIS 180
DB 121 PALGRPPWPLGRASSAKSTPKPGTGPVPSFAKVSDEPPQLEPPQPRGRLOHLOASIS 180
QY 181 QRLGSLDPEMLRCHSEVEDPFGAPKACRPDGSSESOILIGESAVLDPGASGQEPAS 240
DB 181 QRLGSLDPEMLRCHSEVEDPFGAPKACRPDGSSESOILIGESAVLDPGASGQEPAS 240
QY 241 AFOEVSIRVGSPPSSSGGEEKRWNEEPWESPAPVOQESSQAGPPEGAGAVAEEDPPG 300
DB 241 AFOEVSIRVGSPPSSSGGEEKRWNEEPWESPAPVOQESSQAGPPEGAGAVAEEDPPG 300
QY 301 BEVOAOPPOPCSPSPNRYHGLSPSSQARAKAEGTAPLHIPPRLANDRGNVRLNMKQ 360
DB 301 BEVOAOPPOPCSPSPNRYHGLSPSSQARAKAEGTAPLHIPPRLANDRGNVRLNMKQ 360
QY 361 KHYVGRARLSRLRKQAKWKOKKKKGCGGCGATVTTKESCFLEQGDHMAAQCPRPA 420
DB 361 KHYVGRARLSRLRKQAKWKOKKKKGCGGCGATVTTKESCFLEQGDHMAAQCPRPA 420
QY 421 SEEDTDVAGPEPLVPSPPQVPEVPSLDPTVLPYLSIGPSGQLAETPAEYFOALEQGHQA 480
DB 421 SEEDTDVAGPEPLVPSPPQVPEVPSLDPTVLPYLSIGPSGQLAETPAEYFOALEQGHQA 480
QY 481 FRPGQERAYMRILISGISTLLVPTGAGKSLCTYOLPALLYSRSPCLTLVSPILSLMDQ 540
DB 481 FRPGQERAYMRILISGISTLLVPTGAGKSLCTYOLPALLYSRSPCLTLVSPILSLMDQ 540
QY 541 VSGLPPECLKA 550
DB 541 VSGLPPECLKA 550

RESULT 10
ABB63317
ID ABB63317 standard; protein; 1579 AA.
XX
AC ABB63317;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 16743.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.

XX
PD 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL07420.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.

XX Disclosure; SEQ ID NO 16743; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signaling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
ABBS7072). The sequence data for this patent did not form part of the
printed specification, but was obtained in electronic format directly
from WIPO at ftp.wipo.int/pub/published_pcc_sequences

XX Sequence 1579 AA;

Query Match 26.8%; Score 1723.5; DB 4; Length 1579;
Best Local Similarity 28.9%; Pred. No. 1.1e-116;
Matches 467; Conservative 217; Mismatches 459; Indels 471; Gaps 45;

QY 10 RLQAWERAFRRGRGRRSDVAAARETALREYTLRTTQAGGGR 51
DB 10 RLQAWERAFRRGRGRRSDVAAARETALREYTLRTTQAGGGR 51
QY 16 RYKVEKDPKKKNGVSPSKYDIRDASQELRDSYKMYKLTSTFLBETLNDVLSDDGYDL 75
DB 16 RYKVEKDPKKKNGVSPSKYDIRDASQELRDSYKMYKLTSTFLBETLNDVLSDDGYDL 75
QY 52 -TGQAGG-----GIRSS---ESLP----- 66
DB 52 -TGQAGG-----GIRSS---ESLP----- 66
QY 76 EMSQASDFGVSMIDVSLNBSGPOLPLDISALVGPQSSGNLEIPQSVGSEFSNLIDLPN 135
DB 76 EMSQASDFGVSMIDVSLNBSGPOLPLDISALVGPQSSGNLEIPQSVGSEFSNLIDLPN 135
QY 67 -----AAAEAP-BPRCWPMLNRAATKSPPTPGRSRGSVDYQRLKANKTLOAG 100
DB 67 -----AAAEAP-BPRCWPMLNRAATKSPPTPGRSRGSVDYQRLKANKTLOAG 100
QY 136 ROYLTNLVNRDENHYIRKEFAVEELPYNONAMGLVSKPPAPPOVBAKSAFCHGKO- 194
DB 136 ROYLTNLVNRDENHYIRKEFAVEELPYNONAMGLVSKPPAPPOVBAKSAFCHGKO- 194
QY 101 SVPDYQRLKANKTLOAGPALGR--PWPGLGRASSAKSTPKPGTGPVPSFAKVS 157
DB 101 SVPDYQRLKANKTLOAGPALGR--PWPGLGRASSAKSTPKPGTGPVPSFAKVS 157
QY 195 --PKAGASLKPSLSAKLFGSSRGPAKRNPRKRLSCVSSSTSTLSVPTDHHBELDF 252
DB 195 --PKAGASLKPSLSAKLFGSSRGPAKRNPRKRLSCVSSSTSTLSVPTDHHBELDF 252
QY 158 EPPQLEPPQPRGRLOHLOASISGLSDPGLOR-----CHSEV- 199
DB 158 EPPQLEPPQPRGRLOHLOASISGLSDPGLOR-----CHSEV- 199
QY 253 ETILIRKQOEYKXKQOAIANNPMLASHSKESIKT---LVDDGWLRRNTKENTLDEVEVF 309
DB 253 ETILIRKQOEYKXKQOAIANNPMLASHSKESIKT---LVDDGWLRRNTKENTLDEVEVF 309
QY 200 -----DFLAGPK-----ACRPDGSSESOILIGESAVL-----GP 230
DB 200 -----DFLAGPK-----ACRPDGSSESOILIGESAVL-----GP 230
QY 310 AEAANNNGSTSKTTFNGLANLIDLSKLPYKKEKVTQAKPDQWALIOELOTDNNSMNQKP 369
DB 310 AEAANNNGSTSKTTFNGLANLIDLSKLPYKKEKVTQAKPDQWALIOELOTDNNSMNQKP 369
QY 231 GAGSQGEASAFQOVSIRVGSPPSS-----SGGEKRRWNE----- 267
DB 231 GAGSQGEASAFQOVSIRVGSPPSS-----SGGEKRRWNE----- 267
QY 370 DHIINTPPASSQKSAVAPK-NKPPSEQETDSDSVVASSEEQEQRQVRSKRRKIYS 428
DB 370 DHIINTPPASSQKSAVAPK-NKPPSEQETDSDSVVASSEEQEQRQVRSKRRKIYS 428
QY 268 -----PWESEPAVOQES--SQAGP----- 284
DB 268 -----PWESEPAVOQES--SQAGP----- 284


```

Db 429 TASGKVEAFAVEIPNKVPEPETETFAQENPDFSADDEQDATYVVENKKDKAKKQAAOK 488
Qy 285 -----PSEGAGAVAEDEDPGEVQAQPPPP----- 310
Db 489 OKTKPKAEKPKTEKKAKYKAEKKPKAEKKPRNSKKALAVEAPAPPEDEQPLMPEDL 548
Qy 311 -----CSSPSNPRYH-----GLSPSQARAGKAEGTAP 338
Db 549 KYVLLALEAGDITSVRINQDLEADATQRYITRTFAAPNGELSGSINRDEKRAAAR 608
Qy 339 LHPRLAARD-RGNVYRLMKQKHYRG-RALRSRLRKQAWKQMR-----KKG 387
Db 609 KLEERIAAGKLNENVTINIOKKKPKRGKSVNFSKYKQQRHKKRVAALSGPMDMG 668
Qy 388 EFGGGGATVTTKESCFLEQFDHMAAC-----PRP----- 419
Db 669 GCDGG---VL---CFQCGVGHFAQCKVKGDSLPLPSAQLSEDPSPPTLAQENWA 721
Qy 420 -----ASEEDTDAVGPPEPLVPSPOP----- 439
Db 722 SGGAIVVAHSRNTSLPQANAAALQODELNESEDEQSSGDEBVQGHDPNWSDEMDV 781
Qy 440 -----VPE-----VPSLDPT-----VL 451
Db 782 DEBALDAVAEASLSQVPSQEKASPIKTYVGHKIPPEFLKQAGLDTTASSNRSGHGVK 841
Qy 452 PLYSGSPGQLAETPAEVQALFQGHQAFPRQGEAAVNRILISGISTLLVLTGAGKSLC 511
Db 842 PLYDLIPDSVODTTPPEVLEALHMFQHTNFRKQODAIRKTLISGLSISLTLTSGSKSLC 901
Qy 512 YQPLALYSRSPCTTLVSPPLSLMDQVSGLPCLCAKACIHSQTRKQSRVSLQKIRA 571
Db 902 YQPLALYSRKAATLTLVSPVLSLMDQVTVPHNLRHCHLHNTQTAQKRIQOMIAN 961
Qy 572 AQVHVLMTPEALV-----GAGGLPRAQLPPEVAFACIDEAHCLSQMSHNRPCYLRVC 625
Db 962 GEIDILVSPALVAVGERATFGAL--LQQLPPIAFACIDEAHCLSQMSHNRPSYLMIC 1019
Qy 626 KYLRRMGVHCHGLTATRTRTASDVAGHLAAVE-EPDLHGAPAPPTLHLISVMDRPT 684
Db 1020 KYLRNKLGVRTVGLTATATLPTRVSIINHLGISDSEERIIIDIPDLVLVSVSKDENR 1079
Qy 685 DOALLTLQKRFQNDLSIIYCNRRREDTERIALLRTCLHAAWVSGSGRAPKT---T 740
Db 1080 DAALLQNLNSERFECQSIILICTRDECEERLAGFRTVQDRREPTQOTKKRKAVNMQ 1139
Qy 741 AEAVHAGMCSRRERVRORAFMOGOLRVVAVTAFQGLDRPVRATLHGLRPSESVYQ 800
Db 1140 AEPYHAGMPASRRRTVQKAFMSNELRIYVATIAFGKINKPDIRAVIHYNMPRNFSYVQ 1199
Qy 801 AVGRAGRGQCPAHCHLFLQPOGEDLRELRHYHADSTDPLAVKRLVQVFPACTCTTRP 860
Db 1200 EIGRAGRGGLPSHCHFLDAKAGDOSELRRHYSNSIDHVRKLOKLFVCSJC----- 1254
Qy 861 PSEOGAVGGERPVKPPQGEAQLSHQAAPGR-RVCMGHERALPIQTLVQALMPREBA 919
Db 1255 --DKASAKRTALPIP-----LEGDPFRVHMGCHIGTSVEKTEVMKLDIPARN 1300
Qy 920 IETLLCYELAHPHMLBELATTYTHCRNLCPGSPALQALAHRCPLAVCLAAQOPEDEP 978
Db 1301 ISTLLCYMBELDRWCISVLSAYVMAKVISYCGPKYLRKHAKECPPLAMALALQIRDKTF 1360
Qy 979 GGGSSSVFEDMKVLVDSMGWELASVRALCOQMDHEPTGVRRGIGVIVSESLAFHNR 1038
Db 1361 KEDSNVIESVVDIAAGIGMNSGVVYKQDKLEW--VKVNGAPKSPITVSEYDGLFRK 1418
Qy 1039 SPGDLTAEBKQICDPLVGRVORAROLARLRRTQFAHVAFAPSCGPCLCQO--DEERS 1097
Db 1419 VPGDTSEIDNALDILYTRSKQERTQILQLOVYAHGLAAVAYSSCGCCNADFPQDRG 1478
Qy 1098 TRLKDLGRYFEESGQ-----EPQGMEDAQPEFGQARLQDWEQVACDIQRLSLNR 1151
Db 1479 EQLKAIIVRYFANDYFQDLELEIPSNVDP-----ENIIDVHALINMP 1523

```

```

Qy 1152 EEKSSRAVARIFHGIGSPCAQVYQDRREWRKYLHLSFALVGLATEELLQ 1205
Db 1524 DNTFTGRNRIARIFHGIMSPYAVIWR-CKEVRHVKVDENRILHLAMALIK 1576

```

```

RESULT 11
ID AUJ34821
ID AUJ34821 standard; protein; 610 AA.
XX
AC AUJ34821;
XX
DT 14-FEB-2002 (first entry)
XX
DE E: coli cellular proliferation protein #402.
XX
KW antisense; prokaryotic cellular proliferation protein; antibiotic;
XX antibacterial; drug design.
XX
OS Escherichia coli.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
21-MAR-2001; 2001WO-US009180.
XX
PR 21-MAR-2000; 2000US-0191078P.
XX
PR 23-MAY-2000; 2000US-0206848P.
XX
PR 26-MAY-2000; 2000US-0207727P.
XX
PR 23-OCT-2000; 2000US-0242578P.
XX
PR 27-NOV-2000; 2000US-0251625P.
XX
PR 22-DEC-2000; 2000US-0257931P.
XX
PR 16-FEB-2001; 2001US-0269308P.
XX
PA (ELIT- ) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zykkind JW, Wall D, Trawick JD, Carr GT;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
XX
DR N-PSDB; AAS52680.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.
XX
PS Example 3; SEQ ID NO 10414; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes
CC themselves and the encoded proteins. The prokaryotes used are Escherichia
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence represents an essential prokaryotic
CC cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 610 AA;

```

```

Query Match 9.8%; Score 631.5; DB 4; Length 610;
Best Local Similarity 37.6%; Pred. No. 3.1e-37;
Matches 164; Conservative 63; Mismatches 134; Indels 75; Gaps 15;

```

```

Qy 474 BOLGQAFRPPGGERAVMRILSGISTLLVPTGAGSLCYQLPALYRSTRPCLTLVVSPL 533
Db 21 EFTFGYQGRPPGGERBIIDVLSGRDCLVWPTGGSLCYQIALLNG-----LTVVVSPL 76
Qy 534 LSLMDQVSGLPP-CLKACIHSMTKQRESVLQKIRAAQVVMLTPEALVAGGLPP 592
Db 77 ISLMDQVDOQLQANGVAAACINSTQREQOLEVMGCRGTQIRLILYIABERIMLDNFIH 136
Qy 593 AAQLPPVAFACIDEAHCSOMSHNRPCTLYRCVXLRBMGNHCFGLTATRTTASV 652
Db 137 LAHNPVLLA-VDEAHCSIQMGHDFRPEYALGQ-LRQRFPTLPFALTAATADTTRODI 194
Qy 653 ACHLAFAEBPDHGPAPVPTNLHLSVMDRDTQALLTLQSKRFQNLDSI----- 703
Db 195 VRLLG-----LNDPL-----IQIS-SFPRPIRMYLM-----EKKRPDQLMRVYQEQNG 238
Qy 704 ---IYCNREDTERIAALLRTCLHAAMVPSGGRAPKTTAAYAHAGMCSRERRRYORAF 760
Db 239 KSGIYCNRSRAKVEDTAARL-----OSKISAAAYHAGLENNYRAVYQEKF 284
Qy 761 MGGQLRVVAVATVAFPMGGLDRPDVRAVHLGLPSPESYQVANGRAGDQPAHCHLFL-- 818
Db 285 QKDDLIQIVAVATVAFPMGINKPNVRFVHFDIFRNIESYQETGRAGDGLPEAMLFYDP 344
Qy 819 -----QPGEDLRELRH--VHADSTDFLAVKRLV-----QKVPACTC- 855
Db 345 ADMAMLRCLBEKPGQ-Q-LQDIERHKLAMGAFAPAGQTCRLVLANYPBEGQBPQGNCD 403
Qy 856 TCTRPPEQEGAVGE 871
Db 404 ICLDEPKQYDGSIDAQ 419

```

RESULT 12

```

ABU15433
ID ABU15433 standard; protein; 610 AA.
AC ABU15433;
DE 19-JUN-2003 (first entry)
DE Protein encoded by Prokaryotic essential gene #960.
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
OS Escherichia coli.
XX WO200277183-A2.
XX 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
XX 25-OCT-2001; 2001US-0342923P.
XX 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362699P.
XX (ELIT-7) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW,
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Foreysch RA, Xu HH;
XX DR WPI; 2003-029926/02.
XX N-PSDB; ACA19303.
XX New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX Claim 25; SEQ ID NO 43357; 1766bp; English.

```

CC The invention relates to an isolated nucleic acid comprising any one of
CC the 623 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 610 AA;

```

Query Match          9.8%; Score 631.5; DB 6; Length 610;
Best Local Similarity 37.6%; Pred. No. 3,1e-37;
Matches 164; Conservative 63; Mismatches 134; Indels 75; Gaps 15;

Qy 474 BOLGQAFRPPGGERAVMRILSGISTLLVPTGAGSLCYQLPALYRSTRPCLTLVVSPL 533
Db 21 EFTFGYQGRPPGGERBIIDVLSGRDCLVWPTGGSLCYQIALLNG-----LTVVVSPL 76
Qy 534 LSLMDQVSGLPP-CLKACIHSMTKQRESVLQKIRAAQVVMLTPEALVAGGLPP 592
Db 77 ISLMDQVDOQLQANGVAAACINSTQREQOLEVMGCRGTQIRLILYIABERIMLDNFIH 136
Qy 593 AAQLPPVAFACIDEAHCSOMSHNRPCTLYRCVXLRBMGNHCFGLTATRTTASV 652
Db 137 LAHNPVLLA-VDEAHCSIQMGHDFRPEYALGQ-LRQRFPTLPFALTAATADTTRODI 194
Qy 653 ACHLAFAEBPDHGPAPVPTNLHLSVMDRDTQALLTLQSKRFQNLDSI----- 703
Db 195 VRLLG-----LNDPL-----IQIS-SFPRPIRMYLM-----EKKRPDQLMRVYQEQNG 238
Qy 704 ---IYCNREDTERIAALLRTCLHAAMVPSGGRAPKTTAAYAHAGMCSRERRRYORAF 760
Db 239 KSGIYCNRSRAKVEDTAARL-----OSKISAAAYHAGLENNYRAVYQEKF 284
Qy 761 MGGQLRVVAVATVAFPMGGLDRPDVRAVHLGLPSPESYQVANGRAGDQPAHCHLFL-- 818
Db 285 QKDDLIQIVAVATVAFPMGINKPNVRFVHFDIFRNIESYQETGRAGDGLPEAMLFYDP 344
Qy 819 -----QPGEDLRELRH--VHADSTDFLAVKRLV-----QKVPACTC- 855
Db 345 ADMAMLRCLBEKPGQ-Q-LQDIERHKLAMGAFAPAGQTCRLVLANYPBEGQBPQGNCD 403
Qy 856 TCTRPPEQEGAVGE 871
Db 404 ICLDEPKQYDGSIDAQ 419

```

RESULT 13

ABU28286

ID ABU28266 standard; protein; 609 AA.
 XX
 AC ABU28266;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein encoded by Prokaryotic essential gene #13813.
 XX
 KM Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX
 OS Enterobacter cloacae.
 XX
 PN WO20027183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 XX
 PR 06-SEP-2001; 2001US-00948993.
 XX
 PR 25-OCT-2001; 2001US-0342923P.
 XX
 PR 08-FEB-2002; 2002US-00072851.
 XX
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyckind JW,
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 DR MPI: 2003-029926/02.
 XX
 PS N-PSDB; ACA512156.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 25; SEQ ID NO 56210; 1766bp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pcc_sequences
 CC
 CC Sequence 609 AA;
 CC
 CC Query Match 9.6%; Score 619; DB 6; Length 609;

Best Local Similarity 35.4%; Pred. No. 2,6e-36;
 Matches 169; Conservative 73; Mismatches 152; Indels 84; Gaps 18;
 QY 474 EQLGHQAFRRPGQERAVMRILSGISTLLVLPFGAGSLCYQLPALYRSRPPCLTLVSGLP 533
 DB 19 ERFYGQAFRRPGQETIIEIVLSEGRDCLVMPFGGKSLCYQVAVLVING----LTVVSP 74
 QY 534 ISLMDQVSG--LPPCLKAACHSGMTRKORSEVLKIRAAQVHVLMPREALVGAAGLPP 592
 DB 75 ISLMDQVQDLQLANGVAACLNSTQIREQOQVEMGACRTGQRLMTIABERLMDNFUDH 134
 QY 593 AAQLPPVAFACIDEACLSQMSHNRPCYLRYCKYLREMGVHCPLGTATATRTASDV 652
 DB 135 LAHNPVLLA-VDEAHCSIQWGHDFRPEYALGQ-LRQRPFLPMAVLATADDTTRLDI 192
 QY 653 AQHLVAEERDLHGAPVPTNLHLSV-SMDRDTQALTLLOGKRPQNDST----- 703
 DB 193 VRLLG-----LNDP-----YIQVSSFRDNIRYMLA---EKFKPLDQLLRVQEOR 235
 QY 704 ----IYYCRRREDTERIALALTCTCLAAAVPSSGGRAPKTTAAVHAGCSERRRQORA 759
 DB 236 GKSGLIYYCNSRAKVEDTARLQN-----KGFSAAYHAGLENIITRADVQEK 281
 QY 760 FMQGLRVVATVAFGMDLRPDVRAVLHGLPPSFESYQAVGRAGRDQPAHGLFL- 818
 DB 282 FQRDLQIVATVAFGMDGINKPNVRFVHFDIPRNIESYQVETGRAGRDGLPAEAMLPVD 341
 QY 819 -----QPGEDLRELRLH--VHADSTDFLAVKLV-----QRPVPACTC 855
 DB 342 PADMAMLRRCLEKEFGQ-LQDIERHKLNAMGAPAEQTCRLVLVNYFGESRQECGNC 400
 QY 856 -TCTRPSPQEGAVGGERVPRK-YPPQEAQSLSH-----QAAPGRVPCMGHERALPI 906
 DB 401 DICLPBPQYDGLMDARKALSTTYVNOQPFMGVVEVLRGANORIRDMGHDK-LPV 457
 RESULT 14
 ID ABU47859
 ABU47859 standard; protein; 609 AA.
 XX
 AC ABU47859;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein encoded by Prokaryotic essential gene #3386.
 XX
 KM Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX
 OS *Salmonella typhi*.
 XX
 PN WO20027183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 XX
 PR 06-SEP-2001; 2001US-00948993.
 XX
 PR 25-OCT-2001; 2001US-0342923P.
 XX
 PR 08-FEB-2002; 2002US-00072851.
 XX
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyckind JW,
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 DR MPI: 2003-029926/02.
 XX
 PS N-PSDB; ACA51729.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 75783; 1766bp; English.

CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 609 AA;

Query Match 9.5%; Score 609.5; DB 6; Length 609;
 Best Local Similarity 37.7%; Pred. No. 1.3e-35;
 Matches 162; Conservative 58; Mismatches 137; Indels 73; Gaps 14;

QY 474 BOLGHOAFRPGGERAVMRILSGITSLVLPPTAGKSLCYQLPALYSSRSPCLTIVVSP 533
 DB 19 EFTFGYQGFPGQBAIITDIALSGRDCLVWPTGGGSLCYQIPALLDG---LTVVVSPL 74
 QY 534 LSLMDVDSG-LPPLCKAACHSGMTRKQRESVLQKIRAAQVHVIMLTPEALVAGGLPP 592
 DB 75 LSLMDQVQLANGVAAACLNSTQSRQOLEVMACGRTGQIRLLYTABRMLDNFLDH 134
 QY 593 AAOLEPVAFACIDEAHCLSQWSHNPFCYLRVCKVLBERMGVCFGLTATATRTYASDV 652
 DB 135 LHMNPVLLA-VDEANICISQMGHDPPEYALAGQ-LRQRPALPFAVLTATDTRRQI 192
 QY 653 AOHLLVAEPPDLHGPAVPTNLHLSVSMRDTDQALLTLQGRFONLDSI----- 703
 DB 193 IRLG-----LNDPL-----IQIS-SPDRPNIRVYLM-----EKFPDLQLMRYVOERG 236
 QY 704 ----IYCNREDEPTERIAALLRTCLHAAMPVSGSGRAPKTTABVYHMGSRERRRQAP 760
 DB 237 KSGIITCNRAKVEDTPARIQ-----SRG-----ISAAVYHGLENAIADVQEK 282
 QY 761 MGOQLRVVAVTAVFGMLDRPDVRAVLTGLPPSPFSYVOAVGRAGRDQPAHCHFLDP 820
 DB 283 QRDDLIQIVAVTAVFGMGINKPNRVFVHHDPINISYVOETGRARDGIPAEAMLPYDP 342
 QY 821 -----QSEDLRELRRHVHADSTDFLAARLV-----QRVFPACTC-T 856
 DB 343 ADMAMLRCLBEKPAQQLDIERHKINAMGAFAEAQTCRLVLTANFYGEGRQPCNCNDI 402
 QY 857 CTRPPEOEG 866
 DB 403 CLDPPKQYDG 412

RESULT 15

AAU38154
 ID AAU38154 standard; protein; 615 AA.

AC AAU38154;

DT 14-FEB-2002 (first entry)

DE Salmonella typhi cellular proliferation protein #45.

KW Antisense; prokaryotic cellular proliferation protein; antibiotic;
 KW antibacterial; drug design.

OS Salmonella typhi.

PN WO200170955-A2.

PD 27-SEP-2001.

PF 21-MAR-2001; 2001WO-US009180.

PR 21-MAR-2000; 2000US-0191078P.

PR 23-MAY-2000; 2000US-0206848P.

PR 26-MAY-2000; 2000US-0207727P.

PR 23-OCT-2000; 2000US-0242578P.

PR 27-NOV-2000; 2000US-0253625P.

PR 22-DEC-2000; 2000US-0257931P.

PR 16-FEB-2001; 2001US-0269308P.

PA (ELIT-) ELITRA PHARM INC.

PI Haselbeck R, Ohlssen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;

DR WPI, 2001-611495/70.

DR N-PSDB; AAS56013.

PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids.

XX Example 3; SEQ ID NO 13747; 511bp; English.

XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the genes,
 CC their use in the discovery of novel antibiotics, the essential genes
 CC themselves and the encoded proteins. The prokaryotes used are *Escherichia*
 CC *coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*,
 CC *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also
 CC useful for the identification of potential new targets for antibiotic
 CC development. The antisense nucleic acids can also be used to identify
 CC proteins used in proliferation, to express these proteins, and to obtain
 CC antibodies capable of binding to the expressed proteins. The proteins can
 CC be used to screen compounds in rational drug discovery programmes. The
 CC antisense nucleic acid sequence is also useful to screen for homologous
 CC nucleic acids which are required for cell proliferation in a wide variety
 CC of organisms. The present sequence represents an essential prokaryotic
 CC cellular proliferation protein. Note: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 615 AA;

Query Match 9.5%; Score 609.5; DB 4; Length 615;
 Best Local Similarity 37.7%; Pred. No. 1.3e-35;
 Matches 162; Conservative 58; Mismatches 137; Indels 73; Gaps 14;

QY 474 BOLGHOAFRPGGERAVMRILSGITSLVLPPTAGKSLCYQLPALYSSRSPCLTIVVSP 533
 DB 25 EFTFGYQGFPGQBAIITDIALSGRDCLVWPTGGGSLCYQIPALLDG---LTVVVSPL 80
 QY 534 LSLMDVDSG-LPPLCKAACHSGMTRKQRESVLQKIRAAQVHVIMLTPEALVAGGLPP 592

```
Db      81 ISLMEDQVDOQLANGVAAACLNSTQSREBQLEVMACRGTQIRLXIAPERLMDNFDLH 140
OY      593 AAOLPPVAFACIDEAHCLSQWSHNFRPCYLKVCYLREBMGVHCEGLTATATRTASDV 652
Db      141 LAHNPVLLA-VDEAHCTISQWGHDFRPEYALGQ-LRGRFPALPFALTATADFTTRQDI 198
OY      653 AQHLVAEERPDHGPAPVPTNHLHSVSMRDTDQALLTLQGRFQNLDSI----- 703
Db      199 IRLIG-----LNDPL-----IQIS-SFDRPNIRYMLM---EKFRPLDQLMRYVQEORG 242
OY      704 ---IYCNRREBDETRIALLRFTCLHAAMVPGSGGRAPKTTABAYHAGMCSRERRRVQRAF 760
Db      243 KSGIITCNSRAVEDTAAALQ-----SRG-----ISAAYHAGLENAIRADYQEKF 288
OY      761 MOGOLRVVAVATVAFSGMLDRPDVRAVLHLGLPPSFESYVQAVGRAGRDGQPAHCHLFIQF 820
Db      289 QRDLOIVAVATVAFSGMGINKPNVRFVHFDIPNIESYYQETGRAGRDGLPAEAMLFYDP 348
OY      821 -----QGEDLRELRRHVADSTDPLAVKRLV-----QRVPACTC-T 856
Db      349 ADMAMLRRCLEKEPAGQLODIERHKLNAMGAFAPAEQTCRRLVLNYPGEGROEPCGNCDI 408
OY      857 CTRPPEOEG 866
Db      409 CLDPPKQYDG 418
```

Search completed: March 19, 2005, 01:04:38
Job time : 190 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 19, 2005, 01:01:38 ; Search time 165 Seconds
(without alignments)
2419.962 Million cell updates/sec

Title: US-09-889-325-4

Perfect score: 6424
Sequence: 1 MERLADVNERLOAMERARRR.....HLSFALVGLATRELLQVAR 1208

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1401741 seqs, 330541175 residues

Total number of hits satisfying chosen parameters: 1401741

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6424	100.0	1208	US-09-889-325-4	Sequence 4, Appli
2	765	11.9	874	US-10-437-963-121617	Sequence 121617,
3	632.5	9.8	607	US-10-374-077-75	Sequence 75, Appl
4	631.5	9.8	610	US-09-815-242-10414	Sequence 10414, A
5	631.5	9.8	610	US-10-282-122A-43357	Sequence 43357, A
6	621	9.7	370	US-09-753-143-77	Sequence 77, Appl
7	619	9.6	609	US-10-282-122A-56210	Sequence 56210, A
8	609.5	9.5	609	US-10-282-122A-75783	Sequence 75783, A
9	609.5	9.5	615	US-09-815-242-13747	Sequence 13747, A
10	607.5	9.5	615	US-10-282-122A-74973	Sequence 74973, A
11	606	9.4	608	US-10-282-122A-59929	Sequence 59929, A
12	593	9.2	610	US-10-282-122A-78025	Sequence 78025, A
13	589.5	9.2	600	US-10-282-122A-69214	Sequence 69214, A

14	581.5	9.1	620	US-10-282-122A-76978	Sequence 76978, A
15	578.5	9.0	632	US-10-282-122A-67293	Sequence 67293, A
16	578	9.0	619	US-09-815-242-11085	Sequence 11085, A
17	578	9.0	619	US-10-282-122A-58271	Sequence 58271, A
18	554.5	8.6	589	US-10-282-122A-42447	Sequence 42447, A
19	552	8.6	590	US-10-282-122A-60691	Sequence 60691, A
20	549	8.5	597	US-10-282-122A-72864	Sequence 72864, A
21	547.5	8.5	589	US-09-815-242-10803	Sequence 10803, A
22	545	8.5	580	US-09-815-242-4959	Sequence 4959, A
23	545	8.5	712	US-09-815-242-11915	Sequence 11915, A
24	545	8.5	712	US-10-282-122A-66476	Sequence 66476, A
25	544.5	8.5	390	US-10-282-122A-61441	Sequence 61441, A
26	544	8.5	341	US-10-369-493-8990	Sequence 8990, Ap
27	542.5	8.4	1955	US-10-293-604-3	Sequence 3, Appli
28	541	8.4	588	US-10-437-963-141345	Sequence 141345,
29	528	8.2	590	US-10-282-122A-57618	Sequence 57618, A
30	526	8.2	592	US-10-282-122A-70676	Sequence 70676, A
31	525.5	8.2	1328	US-10-374-077-76	Sequence 76, Appl
32	524.5	8.2	658	US-10-156-761-12385	Sequence 12385, A
33	518.5	8.1	715	US-10-282-122A-67821	Sequence 67821, A
34	517	8.0	653	US-10-425-114-59459	Sequence 59459, A
35	515	8.0	827	US-10-282-122A-53151	Sequence 53151, A
36	514	8.0	593	US-10-282-122A-70264	Sequence 70264, A
37	513	8.0	593	US-09-815-242-12306	Sequence 12306, A
38	511	8.0	601	US-10-282-122A-48328	Sequence 48328, A
39	510	7.9	592	US-10-282-122A-71773	Sequence 71773, A
40	508	7.9	335	US-10-369-493-10345	Sequence 10345, A
41	507.5	7.9	720	US-10-282-122A-48551	Sequence 48551, A
42	506.5	7.9	410	US-10-443-108-2	Sequence 2, Appli
43	505.5	7.9	478	US-10-369-493-2682	Sequence 2682, Ap
44	505	7.9	473	US-10-425-114-42472	Sequence 42472, A
45	504	7.8	648	US-10-282-122A-51285	Sequence 51285, A

ALIGNMENTS

RESULT 1
US-09-889-325-4
; Sequence 4, Application US/09889325
; Publication No. US20040224312A1
; GENERAL INFORMATION:
; APPLICANT: AGENE Research Institute, Co., Ltd.
; TITLE OF INVENTION: Gene causative of Rothmund-Thomson syndrome
; FILE REFERENCE: AI-003PCT
; CURRENT APPLICATION NUMBER: US/09/889,325
; CURRENT FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: JP 1999-11218
; PRIOR FILING DATE: 1999-01-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1208
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-889-325-4

Query Match	100.0%	Score 6424	DB 11	Length 1208
Best Local Similarity	100.0%	Pred. No. 0		
Matches 1208	Conservative	0	Mismatches	0
			Indels	0
			Gaps	0
QY	1	MERLADVNERLOAMERARRRGRRRPSDDVEAAPEETRALYREYRTIKRTTGAAGGLR	60	
DB	1	MERLADVNERLOAMERARRRGRRRPSDDVEAAPEETRALYREYRTIKRTTGAAGGLR	60	
QY	61	SSSESLPAAEAEPEPCRGPHLNRAATKSPPTPGRSQSVDPYGGRLKNTLKTLOAG	120	
DB	61	SSSESLPAAEAEPEPCRGPHLNRAATKSPPTPGRSQSVDPYGGRLKNTLKTLOAG	120	
QY	121	PALGRPPWPLGRASSKASTPKPGTGPVPSFAEKVSDPPLPBPQPPGRLQHLQASLS	180	
DB	121	PALGRPPWPLGRASSKASTPKPGTGPVPSFAEKVSDPPLPBPQPPGRLQHLQASLS	180	

181 ORLSGSLDPGMLORCHSEVDFPLGAPKACRPDLGSEESQULLIPGESAVLPGAGSQGPBAS 240
181 QRLGSLDPGMLORCHSEVDFPLGAPKACRPDLGSEESQULLIPGESAVLPGAGSQGPBAS 240
QY 241 AFOEVSIRVGSFQPSGSGGKRWNEEWESPAQVOQESSQAGPSEGAQAVAEEDPPG 300
Db 241 AFOEVSIRVGSFQPSGSGGKRWNEEWESPAQVOQESSQAGPSEGAQAVAEEDPPG 300
QY 301 EFWOAPQPCSSPSNPRYHGLSPSSQAGAKAGETAPLHPRLARHNGVYVRLNMQO 360
Db 301 EFWOAPQPCSSPSNPRYHGLSPSSQAGAKAGETAPLHPRLARHNGVYVRLNMQO 360
QY 361 KHYVGRALRSRLRKQAWKOKMKKKGECFGGGATVTTKESCFLENEQPDHMAAQCPRA 420
Db 361 KHYVGRALRSRLRKQAWKOKMKKKGECFGGGATVTTKESCFLENEQPDHMAAQCPRA 420
QY 421 SEEDTDVAGPEPLVPSPOVPEVPSLDPVPLVLSYSGSGQLAETPAEVPQALBOLGHQA 480
Db 421 SEEDTDVAGPEPLVPSPOVPEVPSLDPVPLVLSYSGSGQLAETPAEVPQALBOLGHQA 480
QY 481 PRPGORAVMRILISGISTLLVLTPTGAGKSLCYQLPALVLSRSPCLTLVVSPLSLMDQO 540
Db 481 PRPGORAVMRILISGISTLLVLTPTGAGKSLCYQLPALVLSRSPCLTLVVSPLSLMDQO 540
QY 541 VSGLPPECLKACISHSMTIRKORSEVLOKIRAAQVHVLMTPREALVGAGGLPRAAQLPVA 600
Db 541 VSGLPPECLKACISHSMTIRKORSEVLOKIRAAQVHVLMTPREALVGAGGLPRAAQLPVA 600
QY 601 FACIDBAHCLSQMSHNFRCYLRVCYVLEBNGVHCFGLTATATRTTSDVAQOHLAVAE 660
Db 601 FACIDBAHCLSQMSHNFRCYLRVCYVLEBNGVHCFGLTATATRTTSDVAQOHLAVAE 660
QY 661 EBDLHGPAVPTNLHLSVSMDBRDTQALTLLOQKGFONLDSIIICNREDTERIAALL 720
Db 661 EBDLHGPAVPTNLHLSVSMDBRDTQALTLLOQKGFONLDSIIICNREDTERIAALL 720
QY 721 RFLCHAAMVPGSGGGRAPKTTAAEYHAGMCSREBRVQRAFMQOQLHVVAATVAFGGLDR 780
Db 721 RFLCHAAMVPGSGGGRAPKTTAAEYHAGMCSREBRVQRAFMQOQLHVVAATVAFGGLDR 780
QY 781 PRVRAVTLHGLPSPFSYVOAVGRARDQOPACHFLPOQEGDLRLRRHVADSTDL 840
Db 781 PRVRAVTLHGLPSPFSYVOAVGRARDQOPACHFLPOQEGDLRLRRHVADSTDL 840
QY 841 AVKRLVQRFPACTCTCTRPSPBOGAVGGERVPRKYPROBAQOLSHQAAPGRVCMGH 900
Db 841 AVKRLVQRFPACTCTCTRPSPBOGAVGGERVPRKYPROBAQOLSHQAAPGRVCMGH 900
QY 901 EBRALPIQLTVQALDMBEBAIETLLCYLLELHPHMLELATTYTHCRINCPCGPAQOQALA 960
Db 901 EBRALPIQLTVQALDMBEBAIETLLCYLLELHPHMLELATTYTHCRINCPCGPAQOQALA 960
QY 961 HRCPLAVCLAOQLPEDPGQSSSVFEDVAVKVDNSGMELASVRAALCOLQMDHNERTEV 1020
Db 961 HRCPLAVCLAOQLPEDPGQSSSVFEDVAVKVDNSGMELASVRAALCOLQMDHNERTEV 1020
QY 1021 RRGTVLVEFSELAFLRSPGDLTAEEKQOICDFLYGVQAREQALRLRTPFOAFHSY 1080
Db 1021 RRGTVLVEFSELAFLRSPGDLTAEEKQOICDFLYGVQAREQALRLRTPFOAFHSY 1080
QY 1081 APFSCGPELEQODEESTRLKDLGRYFEEBEGQEGMEDAQPPEGAQRLQDMEDQYR 1140
Db 1081 APFSCGPELEQODEESTRLKDLGRYFEEBEGQEGMEDAQPPEGAQRLQDMEDQYR 1140
QY 1141 CIRIROLSLRPEKSESSAVARIETHGIGSPCYAUYGODRRFWRKYLHLSFHALVGLAT 1200
Db 1141 CIRIROLSLRPEKSESSAVARIETHGIGSPCYAUYGODRRFWRKYLHLSFHALVGLAT 1200
QY 1201 EELLQVAR 1208
Db 1201 EELLQVAR 1208

RESULT 2
US-10-437-963-121617
; Sequence 121617, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Mu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Bardazuk, Brad
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 121617
; LENGTH: 874
; TYPE: PR
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(874)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_24624C.1.pep
US-10-437-963-121617

Query Match 11.9%; Score 765; DB 16; Length 874;
Best Local Similarity 24.1%; Pred. No. 3.2e-41;
Matches 274; Conservative 162; Mismatches 380; Indels 320; Gaps 39;
QY 110 KANLKGTLQAGPAGRRPWRPLGRASSKASTPKRP--GTGVPSPAEKYSDBPOLPEPQ 166
Db 13 EGSLLSDVASAPRRRRSP--PRPAPRRPPRRPKHRRPAPRTKPKKTPRPAASABA 69
QY 167 PRGRLOHQAISLQSLSLD--GWLORCHSEVDFPLGAPKACRPDLGSEES--QLILP 222
Db 70 PAPPPTTIRAAALS--DPHGLAR-----IAAGSLTAASGASSSFRRLVQ 116
QY 223 GESAVLPGAGSQGEASAFQEVSIKVGSPQPSGSGGKRWNEEWESPAQVOQESSQA 282
Db 117 SNBSPDPAATATAASSAPSEVPASAARPPRTAA-----TDAPQTRPRVHP 165
QY 283 GPPSAGAVAEEDPPGEPVQAPPPQPCSSPSNPRYHGLSPSSQAGAKAGETAPLHP 342
Db 166 NSVSEVAAASAAAEOP-----KARGG----- 187
QY 343 PRLAHDRGNVYRLNMKQHYVGRALRSRLRKQAWKOKMKKKGECFGGGATVTTKES 402
Db 188 -----SEGNFRLNI--NGYGRKRTFKSQAKSTKCSWKRQ-----RAAGT----- 229
QY 403 CFLENEQPDHMAAQCPRPASBEDTDVAGPEPLVPSPOVPEVPSLDPVPLVLSYSG 462
Db 230 -----FRSQGDEGDLVAEALLERKQASD-----SVLEAVE-----SV 264
QY 463 AETPAEVPQALBOL-----GHOAFRPGQERAVMRILISGISTLLVLTPTGAGKSLCYQ 517
Db 265 REDPSE--ONLKSLLNAAVGHDSFQOQLEAIQOIVAGESTLVLVPTGAGKSLCYQVPM 322
QY 518 LYSRSPCLTLVVSPLSLMDQOVGLPCLKACISHSMTIRKORSEVLOKIRAAQVHV 577
Db 323 IL-----PGLTLVVSPLSLMDQOLKRLPAFLPGGLASSQTSDEPHDTQLRAAGET 378
QY 578 MLTPREALVGAGGLPRAAQLPVAFACTIDEAHCLSQMSHNFRCYLRV--CKVLRERMGVAC 636
Db 379 FVSPERFLNEERLULFROTLPISTIVAIIDEAHCLSQMSHNFRCYLRSLRLRKLVNQC 438
QY 637 FLGLTATATRTTASDVQHLAVAEEDDLHGPAVPTNLHLSVSMDBRDTQALTLLOQKR 696

PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 10414
LENGTH: 610
TYPE: PRF
ORGANISM: Escherichia coli
US-09-815-242-10414

Query Match 9.8%; Score 631.5; DB 9; Length 610;
Best Local Similarity 37.6%; Pred. No. 1.1e-32;
Matches 164; Conservative 63; Mismatches 134; Indels 75; Gaps 15;

474 BOLGHAAPRQGERAVMRLISGISTLVLPTGAGKSLCYQLPALYXSRSPCLTIVASPL 533
21 EFTFGYQRRPQGEELIDYVLSGRDCLVMPFGGKSLCYQLPALING---LIVVASPL 76
534 ISLMDQVSGLP-CLKACIHSGMTKQRESVLOKIRAAQVHVMTPEALVAGAGLPP 592
77 ISLMDQVQDLOANGVAAACLNSTOTREQOLEVMGTCRTGQIRLYIAPERLMDNPLEH 136
593 AAOLPPVAFACIDEAHCLSQSHNRPCLYKVLKREMGVHCFGLTATATRTASV 652
137 LAHWPVLLA-VDEAHCLSQSHNRPCLYKVLKREMGVHCFGLTATATRTASV 194
653 AOHLAVAEPDLHGAPVPTNLHLSVMDRTDQALLTLQCKRQNDLSI----- 703
195 VLLG-----LNDPL-----IQIS-SFDRPNIRYMLM-----EKFPDLQLRVYQEOGR 238
704 ---IYCNREDTERIALLRCTCLHAAMVPGSGGAPKTTAAVYAGMCSRERRVORAF 760
239 KSGIITCNRAKVEDTARL-----QSKGISAAAYAGLENNVRADVQERF 284
761 MGGQLRVVATVAFMGJLDPRVAVLHGLPPSFESYQAVGRRDQPAHCHLFL-- 818
285 QSDDLQIVATVAFMGJINKPNRVFVHFDIPRNIESYQETGRAGRDGLPAEAMLFYDP 344
819 -----QPOGDLRELRRH---VHADSTDFLAVKRLV-----QVFPACGC- 855
345 ADMAMLRCLERKPPQGLQDIERHKLAMGAFAEAGTCRRLVLTNYGEGQEPGCGND 403
QY 856 TCTRPSEQEGAVGGE 871
DB 404 ICLDPPKQYDGSSTDAQ 419

RESULT 5

US-10-282-122A-43357
Sequence 43357, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Mang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Foreytch, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: Patent version 3.1
SEQ ID NO: 43357
LENGTH: 610
TYPE: PRF
ORGANISM: Escherichia coli
US-10-282-122A-43357

Query Match 9.8%; Score 631.5; DB 15; Length 610;
Best Local Similarity 37.6%; Pred. No. 1.1e-32;
Matches 164; Conservative 63; Mismatches 134; Indels 75; Gaps 15;

474 BOLGHAAPRQGERAVMRLISGISTLVLPTGAGKSLCYQLPALYXSRSPCLTIVASPL 533
21 EFTFGYQRRPQGEELIDYVLSGRDCLVMPFGGKSLCYQLPALING---LIVVASPL 76
534 ISLMDQVSGLP-CLKACIHSGMTKQRESVLOKIRAAQVHVMTPEALVAGAGLPP 592
77 ISLMDQVQDLOANGVAAACLNSTOTREQOLEVMGTCRTGQIRLYIAPERLMDNPLEH 136
593 AAOLPPVAFACIDEAHCLSQSHNRPCLYKVLKREMGVHCFGLTATATRTASV 652
137 LAHWPVLLA-VDEAHCLSQSHNRPCLYKVLKREMGVHCFGLTATATRTASV 194
653 AOHLAVAEPDLHGAPVPTNLHLSVMDRTDQALLTLQCKRQNDLSI----- 703
195 VLLG-----LNDPL-----IQIS-SFDRPNIRYMLM-----EKFPDLQLRVYQEOGR 238
704 ---IYCNREDTERIALLRCTCLHAAMVPGSGGAPKTTAAVYAGMCSRERRVORAF 760
239 KSGIITCNRAKVEDTARL-----QSKGISAAAYAGLENNVRADVQERF 284
761 MGGQLRVVATVAFMGJLDPRVAVLHGLPPSFESYQAVGRRDQPAHCHLFL-- 818
285 QSDDLQIVATVAFMGJINKPNRVFVHFDIPRNIESYQETGRAGRDGLPAEAMLFYDP 344
819 -----QPOGDLRELRRH---VHADSTDFLAVKRLV-----QVFPACGC- 855
345 ADMAMLRCLERKPPQGLQDIERHKLAMGAFAEAGTCRRLVLTNYGEGQEPGCGND 403
QY 856 TCTRPSEQEGAVGGE 871
DB 404 ICLDPPKQYDGSSTDAQ 419

RESULT 6

US-09-753-143-77
Sequence 77, Application US/09753143
Patent No. US20020102550A1
GENERAL INFORMATION:
APPLICANT: NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA

RESULT 8
US-10-282-122A-75783
; Sequence 75783, Application US/10282122A
; Publication No. US20040029129A1

Query Match 9.5%; Score 609.5; DB 15; Length 609;

RESULT 9
US-09-815-242-13747
; Sequence 13747, Application US/09815242
; Patent No. US20020061569A1

TYPE: PRT

```

; ORGANISM: Salmonella typhi
US-09-815-242-13747

Query Match          9.5%; Score 609.5; DB 9; Length 615;
Best Local Similarity 37.7%; Pred. No. 3e-31;
Matches 162; Conservative 58; Mismatches 137; Indels 73; Gaps 14;

Qy 474 BOLGHOAEPGGEARVAMRILSGISTLLVPTGAGSLCYQLPALYSRRSPCLTLVSP 533
    25 EFTFGYQOFRPGQBAIIDTLASGRDCLVWPTGGKSLCYQIPALLDG---LTVVSP 80
Db 534 ISLMDQVSG-LPPLKAACTHSGMTKORRESYLQKIRAOVHVMLTPEALVGAGLPP 592
    81 ISLMDQVQDLANGVAACLNSTOSREQOLEVMAGCRGQIRLYIAERMLDNFLDH 140
Qy 593 AAOLPPVAFACIDEAHCSQSMHNPFCYLRVCYLREMGVHCFGLTATATRTASDV 652
    141 LAHNPVLLA-VDEAHCTISQMGHDFRPEYALGQ-LRQFRPALPFMALATATADTTRODI 198
Qy 653 AOHVAEERPDHGPAPVPTNLHLSVMDRTDQALLTLQGRPNLDST----- 703
    199 IRLG-----LNDPL-----IQIS-SFDRPNIRYMLM---SKKPLDQLMKRYVOEORG 242
Qy 704 ---IYCNREDTERIALLRCTLHAAWPGSGGAPKTTAEYHAGMCSRERRRQVRAF 760
    243 KSGIYCNSRAKVEDTAARLQ-----SRG-----ISAAYHAGLENALRADVQEKF 288
Qy 761 MOGQLRVVATVAFMGGLDRPDVRAVLHLGLPPSFESYQAVGRAGRDQPAHCHLFLDP 820
    289 QRDLDQIVATVAFMGINKPNVRFPVHFDPINIESYVOEGRAGRDPLPAEAMLFYDP 348
Qy 821 -----QGEDLRELRHVHADSTDFLAVKRLV-----QRVPACTC-T 856
    349 ADMAMLRCLBEKPAQGLDIERHKLNMGAFAEAQTCRLVLNYPGEGRQEPGNCIDI 408
Qy 857 CTRPSEBOG 866
    409 CLDPPKQYDG 418
Db

RESULT 10
US-10-282-122A-74973
; Sequence 74973, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Foreyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT FILING DATE: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23

```

```

; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patencin version 3.1
; SEQ ID NO 74973
; LENGTH: 615
; TYPE: PRF
; ORGANISM: Salmonella typhimurium
US-10-282-122A-74973

Query Match          9.5%; Score 607.5; DB 15; Length 615;
Best Local Similarity 37.7%; Pred. No. 4e-31;
Matches 162; Conservative 58; Mismatches 137; Indels 73; Gaps 14;

Qy 474 BOLGHOAEPGGEARVAMRILSGISTLLVPTGAGSLCYQLPALYSRRSPCLTLVSP 533
    25 EFTFGYQOFRPGQBAIIDTLASGRDCLVWPTGGKSLCYQIPALLDG---LTVVSP 80
Db 534 ISLMDQVSG-LPPLKAACTHSGMTKORRESYLQKIRAOVHVMLTPEALVGAGLPP 592
    81 ISLMDQVQDLANGVAACLNSTOSREQOLEVMAGCRGQIRLYIAERMLDNFLDH 140
Qy 593 AAOLPPVAFACIDEAHCSQSMHNPFCYLRVCYLREMGVHCFGLTATATRTASDV 652
    141 LAHNPVLLA-VDEAHCTISQMGHDFRPEYALGQ-LRQFRPALPFMALATATADTTRODI 198
Qy 653 AOHVAEERPDHGPAPVPTNLHLSVMDRTDQALLTLQGRPNLDST----- 703
    199 IRLG-----LNDPL-----IQIS-SFDRPNIRYMLM---SKKPLDQLMKRYVOEORG 242
Qy 704 ---IYCNREDTERIALLRCTLHAAWPGSGGAPKTTAEYHAGMCSRERRRQVRAF 760
    243 KSGIYCNSRAKVEDTAARLQ-----SRG-----ISAAYHAGLENALRADVQEKF 288
Qy 761 MOGQLRVVATVAFMGGLDRPDVRAVLHLGLPPSFESYQAVGRAGRDQPAHCHLFLDP 820
    289 QRDLDQIVATVAFMGINKPNVRFPVHFDPINIESYVOEGRAGRDPLPAEAMLFYDP 348
Qy 821 -----QGEDLRELRHVHADSTDFLAVKRLV-----QRVPACTC-T 856
    349 ADMAMLRCLBEKPAQGLDIERHKLNMGAFAEAQTCRLVLNYPGEGRQEPGNCIDI 408
Qy 857 CTRPSEBOG 866
    409 CLDPPKQYDG 418
Db

RESULT 11
US-10-282-122A-59929
; Sequence 59929, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Foreyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A

```

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OR INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ERIIRA.034
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/240,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27

```

; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-15
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 78025
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Yersinia pestis
US-10-282-122A-78025

Query Match          9.2%; Score 593; DB 15; Length 610;
Best Local Similarity 33.3%; Pred. No. 3.5e-30;
Matches 169; Conservative 71; Mismatches 172; Indels 96; Gaps 18

Cy      474  EQLHQAARPCOEBAVMKRLTSGISITLVLPYGAGSKLTCYQAPALLYSRRSPCLTLVSPPL 533
Db      19  DTFGIQQRPPQOELTNATLTSQDCLVMPFPGGSLCYQAPALV---TDGLTLVSPPL 74
Cy      534  LSLMDQVSGSRPPC-LKACIHSGMTKRSKORSVLOKIRAPQVYVIMLTPREALVGAGLPP 592
Db      75  ISLMKDQVQDLAVGVGAGCINSSQTRBQQLAVMDGCGSSGQIKLILYIAPERLVMESFLDQ 134
Cy      593  AAQLPVAACIDENHCISQSHNRPCYLAVNCKTLRRBMGNHCLGATATRTRTASDY 652
Db      135  LYQWRP-ALLAVDEHNCISQWGHDRPEY-PALGOLKQRPDLRYATLATAIDDEATRGDI 192
Cy      653  AQHLVAVEEPLDGHGPAVPYTNLHLTVSMDRDTQALLTLQGRFQNTDSI----- 703
Db      193  VRLT-----NLDDQL-----IQIS-SFDRIPIRYTLV---EKFPDLQWRFWQDQKG 236
Cy      704  ---IYYCRREDTERIALALLTCLLHAAYVPSGGGAPKPTTAAYHAYAGMSRRRRRYQRAF 760
Db      237  KSGIITCNRSRAKVEDTTRARLQ-----SRG-----LSVAHYAGLNNERRAPQOEAF 282
Cy      761  MGGQLRVVAVATVAFGMLDRPDVRAVLIHLGLPPSFESYVOAGRAGRDQOPAHCHLQRP 820
Db      283  QRDDLQVVAVATVAFGMLGNKPNVRVFVHFDIPRTIESYQETIGRAGRDLPAVAVLLYDP 342

```

QY 821 QG-----EDLRLRRHVHADSTDFLAVKRLVORVF-----PACTC-T 856
DB 343 ADMAMRLRCLERKPAQODIERHKLAMGAPAEBAOTCRRLVILNTFGESKQOPCCNDI 402
QY 857 CTRPSEOGAVGERPVP-KYRPOEADLSH-----QAAPRRVCMGHERALPIQLTV 910
DB 403 CLDPFRKYDGLADAGALSCVYRFGIGYIVEVLRGANNORIRIEMGHDK-----LSV 457
QY 911 QALDMPREAIETLLCYLHPRHMYELL 938
DB 458 YGIGR-----EQTHEWVSVL 473

RESULT 13
US-10-282-122A-69214
Sequence 69214, Application US/10282122A
GENERAL INFORMATION:
APPLICANT: Wang, Lianghu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Foreyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See file Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 69214
LENGTH: 600
TYPE: PRT
ORGANISM: Proteus mirabilis
US-10-282-122A-69214

Query Match 9.2%; Score 589.5; DB 15; Length 600;
Best Local Similarity 32.2%; Pred. No. 5,8e-30;
Matches 184; Conservative 69; Mismatches 151; Indels 167; Gaps 24;

QY 474 BOLGQAPRPGOBRAMRLISGISTLLVYPTGAGSLCYOLPALLYSRSPCLTLIVSVL 533
DB 10 EFTFGGQGFPGGQEIITITITGDCLVNPTGGGSLCYOIPALLUDG---LTVVSVL 65
QY 534 ISLMDQVSGLRPPL---KACIHSQMTKQRESVYLQKIRAAQVHVMLTPEALVAGGL 590

DB 66 ISLMDQVQDL---CLHGIDAAFLNSTQSRDEQLQVQMGORGBIKLYIAPERLMMESFL 123
QY 591 PPAQLPPLVAFACIDBAHCLISQSHNFRPCYLAVCYLBRGVNHCFLGLTATARTTS 650
DB 124 HHLVQWQP-ALLAVDEAHGICISQGHDFRPEY-RGICILROYLPDVBRIALTATADNTTTR 181
QY 651 DVAGHLAVABEPDLHGPAVPNTLHVSQMDRDTQDALTLLOQKRFQNDST----- 703
DB 182 DIINQALNR-----TFL---VHIS-SFDRPNRTYTLV---EKKPLDQIMLFIHQ 225
QY 704 -----IYCNRRREDTERIALRLTCLHAAMVPGSGGARPTAAVAHAGCSRRRRVOR 758
DB 226 KKGSGIYCNRSRKYVETARL-----GKRG-----LSIAVHAAGDITRAKYOD 271
QY 759 AFMOGQLRVVATVAVGSKLDRPDRAVTLHLGLPSESTYQAVGAGDGPACHFL 818
DB 272 AFORDLOIVAVTVAFGKGINKPNVRFVHFPDI PRNIESYUOETGAGBDGPAQAVLFX 331
QY 819 OPGEDRLRLRRHVHADSTDFLAVKRLVORVPACTCTRPSEOGAVGERPVPKYP 878
DB 332 DP--ADMAMLR-----CLDEKPESEOKAI----- 354
QY 879 POEADLSHQAAPRRVCMGHERALPIQLTVQALDMPREAIETLLCYLHPRHMYELL 938
DB 355 --ESHKLAMGAPAEBAOTC---RRL-----VLLNYF----- 380
QY 939 ATTYTHCRNLCPGPAQOLALHRCPLAVCIAQOLPEDPGQSSSVBDMVKLDSMGM 998
DB 381 -----GHRQQA---CGNCDICL-----DPPK---QYD---GLVDA--- 407
QY 999 ELASVRALCOLQWHEPRTGVRGTVLVE 1029
DB 408 -----OKALSCYI---RTGQRFIGIYVE 428

RESULT 14
US-10-282-122A-76978
Sequence 76978, Application US/10282122A
GENERAL INFORMATION:
APPLICANT: Wang, Lianghu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Foreyth, R.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636

```

? PRIOR FILING DATE: 2001-02-09
? PRIOR APPLICATION NUMBER: 60/269,308
? PRIOR FILING DATE: 2001-02-16
? Remaining Prior Application data removed - See File Wrapper or PALM.
? NUMBER OF SEQ ID NOS: 78614
? SOFTWARE: Plasmid version 3.1
? SEQ ID NO 76978
? LENGTH: 620
? TYPE: PR1
? ORGANISM: Vibrio cholerae
? OS-10-282-122A-76978

```

Query Match	9.1%;	Score	581.5;	DB	15;	Length	620;
Best Local Similarity	33.0%;	Pred. NO.	2e-29;				
Matches	160;	Conservative	74;	Mismatches	171;	Indels	80;
						Gaps	14;

QY	456	UGPBGQLEBPAEYFVFOALBEOGHQAPRPGOEBAWRIISGISITLVLPYPGAGKSLCYOYR	515
DB	15	MAESSALPATERPVH--EVFGYQOFRVQOQEVLEALAGRSLVIMPFGGKSLCYQYR	72
QY	516	ALLYSRRSPCLTVVSPPLSLMDDOVSGLP--CLKAACIHSQWTKRKORESVLOKTRAAQV	574
DB	73	ALVLEG-----VTLVISPILSKKDOQDQOLKANGVAEYVNSVTLAREBELLAYVRMHAGQD	128
QY	575	HYLMITPPEALVAGGLPRPAQOQPRPAFACTIDEANCLSQSHNPRFCYLAVCKYLRBRMGV	634
DB	129	KLLVYSPFVRLVAEETIERLSHL--PLAMIAVDEAHCSIQWGHFRFRPYASLQO--LKORFPR	186
QY	635	HCFGLGTLTATRRRTASDVNAQHLAVAEERDLNHRAPRPTLYLHNSVSDMBRDQDALTYLQO	694
DB	187	VPVMAULTATADADATGHDIMQRLQOLNEPHQULGSFDRPNIRYMLVEKHQVSO--VINULET	245
QY	695	KRFQYLDSTIIYCNRRREDTERIALALRTCLDHAWVSGSGRAKTAEAHYHMGSRERR	754
DB	246	ORGQ---CGIYYCSKKEVEMLEKLI--C-----GNBIR-----AASHTAMMDDEBQ	288
QY	755	RVQRAFMOGOLRVVAVATVAFGMDLRDVRVAVLHGLRPSFESUYOAVSGRAGRDGQPAHC	814
DB	289	WVQBAFQRDDLIIVATVAFMGINKPNRVFVWHPDIPNNIESUYOETQGRAGDGLPAEA	348
QY	815	HLFLPOQEDLRELRRHVNADSTDPLAVKLYQKRVPRACSTCTRPSPSEQBAVGGBRRY	874
DB	349	KMLVDP--ADMWMLRRML-----DEKPD 369	
QY	875	PKYPRQEAQOLSHOAPGBRRVY-----MGHERALPLOTLVQALDMP-----EBAI 920	
DB	370	GAQKQVESHKILPAMGAFABAQTCRQVLLNLTGSEYADKFCQGNDCILDRPKRDAIEBAAR	429
QY	921	ETLLC 925	
DB	430	KALSC 434	

```

RESULT 15
US-10-282-122A-67293
: Sequence 67293, Application US/10282122A
: Publication No. US20040029123A1
: GENERAL INFORMATION:
: APPLICANT: Wang, Liangsu
: APPLICANT: Zamudio, Carlos
: APPLICANT: Malone, Cheryl
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Kari
: APPLICANT: Zykkind, Judith
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John
: APPLICANT: Carr, Grant
: APPLICANT: Yamamoto, Robert
: APPLICANT: Forsyth, R.
: APPLICANT: Xu, H.
: TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
: FILE REFERENCE: ELITRA.034A
: CURRENT APPLICATION NUMBER: US/10/282.122A

```

```

; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 67293
; LENGTH: 632
; TYPE: PRT
; ORGANISM: Pasteurella multocida
US-10-282-122A-67293

```

Query Match	9.0%;	Score 578.5;	DB 15;	Length 632;
Best Local Similarity	32.8%;	Pred. No. 3.3e-29;		
Matches 149;	Conservative 76;	Mismatches 142;	Indels 87;	Gaps 15;

```

QY 462 LAEPFAEYFOALBOUHQAFRRGGBERAWRIILSGISTVLVLPETGAGKSLCYOULPALLYSR 521
QY 22 LKQTLADVLHNV -RGYOSFRKGGBEEVIDATILMGXDSITVATIMAGKSLCYOULPALCFEG 79
Db 522 RSPCLTIVVSPILSLMDQVGS -LRPCLKAACIHSGMTERKORESVLOKIRAAOVHVLMT 580
QY 80 ----LTVISPLISLMKQVODQLANGIEADYILNSQFTEBGOOVNKLMSGTLKLLYVS 135
Db 561 PEALVAGAGPPAAOLPVYAPACIDBEAHTLSQWSINFRFCYLAUCKVLAERMGVHCF--- 637
QY 136 PEKWTTSPFHLISIC-KVSPVAIDEAHCTISQMGHDFREYVYQLGSL-----KSCFPHA 188
Db 638 --LGATATRTASTADVNOAHLVAEERPDLCGAPVPYTNLHLSV-SMDRDTQALTLTLQGS 694
Db 189 PIMALTATRADNATRODILRHL-----NLQSP-----HYVIGSDRNRINRTVLV---- 221
QY 695 KRFQNLDSI-----IYCNREDTERIALALFTCLHAANVPQSGGRAEPTTAE 742
Db 232 EKFKEMEDLCRFVLQOKGKGKGIYCNRSKVRIRIESLNN-----KGVSAQ 277
QY 743 AYHAEMCGRERRVORAFMOGOLRVVATVAEGMGJDRDVRVAILHGLPPSESVYQAV 802
Db 278 AYHAELTJSORQVORAFORDNVQVYVATIAFGMGINKSNVAFVYHFDLPJRSIESYYOEL 337
QY 803 GRAGRDGQPAHCHLFLQPG-----EDURELRHHVHADSTDFLAVKRLV 846
Db 338 GRAGRDGDPAAEAVLYPEPADYAMLKILILEKESPORQIEALKLOAIGFAESQTCRRILV 397
QY 847 -----QRVPATC-CCTRPSPSDEAGVGBER 872
Db 398 LNVYGEHQKFCQNCDCICLDPKCKOYDGLTDIAOK 431

```

Search completed: March 19, 2005, 01:12:35
Job time : 169 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 19, 2005, 00:59:02 ; Search time 50 Seconds
(without alignments)
2324.597 Million cell updates/sec

Title: US-09-889-325-4
Perfect score: 6424
Sequence: 1 MERLDVVERLQAWERARR.....HLSFHALVGLATRELLQVAR 1208

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	872	13.6	941	2	A86404
2	631.5	9.8	610	1	BVBCRO
3	631.5	9.8	611	2	F86069
4	631.5	9.8	611	2	H91222
5	609.5	9.5	609	2	AH0917
6	593	9.2	610	2	A10466
7	581.5	9.1	620	2	E82351
8	560.5	8.7	602	2	H97365
9	560.5	8.7	602	2	AH2583
10	552	8.6	590	2	AD1419
11	551	8.6	718	2	AE1832
12	548	8.5	590	2	AE1794
13	545	8.5	712	2	E83226
14	542	8.4	496	2	A69691
15	525.5	8.2	1328	2	S62467
16	523	8.1	824	2	G75413
17	521	8.1	615	2	G87678
18	514	8.0	593	2	B89844
19	513	8.0	592	2	H86850
20	510.5	7.9	676	2	T34609
21	507	7.9	480	2	AB2411
22	506.5	7.9	809	2	T20430
23	506	7.9	1231	2	T24415
24	505.5	7.9	478	2	S77358
25	505.5	7.9	591	2	F69901
26	504	7.8	766	2	H81794
27	499	7.8	491	2	G83850
28	499	7.8	766	2	G81216
29	496.5	7.7	1436	2	T14895

30	493.5	7.7	645	2	A82689	DNA helicase XF138
31	489	7.6	1447	2	S50918	DNA helicase TP51
32	484	7.5	1401	2	T30247	Werner syndrome pr
33	479.5	7.5	1401	2	T17452	Werner syndrome pr
34	478.5	7.4	714	2	G97230	recQ protein, supe
35	476	7.4	665	2	B86243	DNA helicase homol
36	464.5	7.2	659	1	A55311	DNA helicase RECQL
37	463.5	7.2	1417	2	A57570	Bloom's syndrome r
38	458.5	7.1	459	2	B89927	hypothetical prote
39	443	6.9	1031	2	G96634	probable DNA helic
40	437	6.8	467	2	AF1317	ATP-dependent DNA
41	434.5	6.8	467	2	AF1689	ATP-dependent DNA
42	432	6.7	886	2	T16536	hypothetical prote
43	422	6.6	483	2	T51906	related to recQ ge
44	406.5	6.3	607	1	C71367	probable ATP-depen
45	386	6.0	1273	2	E72611	probable ATP-depen

ALIGNMENTS

RESULT 1
A86404
probable protein ATP-dependent DNA helicase RecQ [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #ext_change 31-Mar-2001
C:Accession: A86404
R:Theologis, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lurie, J.S.; Malt, R.; Marshall, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; PMID:21016719; PMID:11130712
A:Accession: A86404
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-941 <STO>
A:Cross-references: GB:AE005172; NID:g10998926; PID:AAG26068.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1
Query Match 13.6%; Score 872; DB 2; Length 941;
Best:Local Similarity 26.2%; Pred.No. 1.4e-38;
Matches 305; Conservative 165; Mismatches 390; Indels 306; Gaps 43;
QY TFGSRQGSVPDYCGRLKANKLTLOAGPALRRPWPGLRASSKASTPRPGTGPVPSFA 152
DB 15 TPRPD---SFPSSPPOLOSFAKRV---PPVSRKM---TSSSSRSKPAAP-THPPNPS 62
QY 93 TFGSRQGSVPDYCGRLKANKLTLOAGPALRRPWPGLRASSKASTPRPGTGPVPSFA 152
DB 15 TPRPD---SFPSSPPOLOSFAKRV---PPVSRKM---TSSSSRSKPAAP-THPPNPS 62
QY 153 EKVSDPEPQLPPROPGRGLQHLQASLSQRLSLDPGMLQRCHEVDPDGLGAPKARPD 212
DB 63 QEPVPPSPYPPPPPPSP-----LFTNLPFRICQSDPA---RSSSVSSP---SRLSRAS 111
QY 213 GSEESQLTIPGESAVLPGAGSQGPESAFQEVISIRVGSPPSSGGEKRNNEBWESE 272
DB 112 FTSVEKLKSDGVDFV-----PEPPLVEVIA----- 136
QY 273 AOYQESSQAGPPSBGAGVAVEEDPPG---BPVQAPQPPQSSPPNRYHGLSPSSQAR 329
DB 137 -----PPK-----SVRRKPPNLTITLITSPPKPVFRSN----- 166
QY 330 AKKAEGTAPLHIFPRLARHDRGVYVLYNKKQGVYVGRALRSRL-----LRQAWKQK 383
DB 167 -GNGE-----GNFYKLNKGR---RKKFPKRYKGVSKRSYSRGR 206
QY 384 RKKGECFGGGAIVTTKESCFINEQFDHMAACPRPASEEDTDVAGPEPLVPSPOCPVPEV 443
DB 207 YKKKADGDSIL--EESDLOKQI-----EDEANGF-----I 238

```

QY 444 PSLDPPTULPLXSLGSPQGLATPBAVFOALQELQHQAFRPGQEAQWNRILISGISTLVLPL 503
Db 239 SSVEDAILLAIVTEASDENLTYKLLNVY-----QYDSFRDQLOAIKMLIGLSSSTMLVLPL 292
QY 504 TGAKSLSICQYCPALVLYSRSPCLTLVWSPLSLMDPDQVSGPLPCLKAACIHSGMTRKQRE 563
Db 293 TGAKSLSICQYCPMML-----PGITLVVSPVLVSMIDQKHLPSIIKGLSSSGRPREAT 348
QY 564 SVLOKIRAAQVHMLMTPEALVGAGGLEPPAAQLPVPAACIDEAHCLSQ----- 612
Db 349 ETLRKLEKGIILKVLFVSPERLLANVEFLSMFMSLSVSLVVDDEAHCVSECIYCLSLCLS 408
QY 613 W-----SHNPPCYLRV--CKYLBERMGHNCGLGATATRTRTASDVAQNLVAEERD 663
Db 409 WLLPAILNRSNHRPSTYKRLKASMLFSELKAECLAMATATYMTLQAVMSLEI--DSTN 467
QY 664 LHGPAPVPTNLHLVSVM---DRODTQALLTLQGRFPONDIIYYCNRREDTERIALYL 720
Db 468 LIQSQJLRDNDELVSLSGANRMD--LLIMESPPYKEINSIIYVCKFYETDMISKYL 525
QY 721 RTCLHAAMVPSGGRAPKPTTAEATHAGCMCSERRRVQAPAFQOGLRVVVAATVAAGMGLDR 780
Db 526 RD-----NNINAKGYHSGPLPADRVRIOESFSCSNIRVVATVAAGMGMDK 571
QY 781 PDVBAVLHLGPSPESPVYQAVAGRAGDGPANCHLFLPQGEDLRELRRHVHNDSTFL 840
Db 572 GDVGAVIHFSVPGSMEBYVQETIGRAGDGRSLSTCHLFF--DNQTYLKLRLASHBDGYDEX 629
QY 841 AVKELVORVPACTCTCTRPPEBQEGAVGGERPVPKYPRQEAQLSHQABPBRVCMGH 900
Db 630 AVGKFLTHVF-----STFKQHE-----KIC-- 650
QY 901 ERALPIQLTVQALMPREAEITLICYLEIHRHMLIELATTYTTCRLNCPGSPAQOLOALA 960
Db 651 --SLVIESASQKPFDMKEEVMQOTILTHLEIGVQYIRMLPQINICCTLN-----F 697
QY 961 HR---CPLAIVCLAQOOLPEDEPGQSGSSVEFPMVYLVDMSG---WELASVRALCOLQMDH 1014
Db 698 HKTSDAQNIIVMLVYQSSPNTLAAARSALVAALIKKXSHVQGLHVDIPRVAASICVAT--- 754
QY 1015 EPRTGVRGRTGVLVFESELAHLNRSPEGDLTAEKD-----QICPFLYGRVOAR 1062
Db 755 -----TVLAEIQAL-----KGEVTEYELKDSAFVCYITLKSPEKICSLSHLTKWL 799
QY 1063 ERQALARLR--TFOAFHSVAFPSGCGPLEQGDDEBRSSTLKQILGRVPEEBEGQEPGMED 1121
Db 800 TEIESCKRKLDIDINSSAAVAAISVSNSELSGAKOTR--SLQSRIFDYFNG-----D 850
QY 1122 AOGEPGQARLODWEQ---VRCDIROFLSLRPEEKFSRAVARIIFHGIGSPCYPAQVY 1177
Db 851 EKCDSPSKA-----TQNCARFLADIKVFLQSNQAKETTPAIRARIMHGVSAPAPVSNW 904
QY 1178 GQDRRFMRKYLHLHSFHALVGLATEEL 1203
Db 905 SK--THFWGRYMWVDFRVIMEAQTEL 929

```

RESULT 2
BVECHO
DNA helicase recQ - *Escherichia coli* (strain K-12)
N:Alternate names: DNA-dependent ATPase recQ (EC 3.6.1.-)
C:Species: *Escherichia coli*
C:Date: 31-Dec-1989 #sequence revision 10-Oct-1997 #text_change 01-Mar-2002
C:Accession: G65186; U00137; J35786; S30712
R:Bibacter, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC
.A.; Rose, D.U.; Mau, B.; Sho, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of *Escherichia coli* K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: G65186
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

```

A:Residues: 1-610 <BLAT>
A:Cross-references: GB:E000458; GB:U00096; NID:G2367299; PION:AACT6825.1; PID:G2367301.
A:Experimental source: strain K-12, Substrain M01655
R:Ritino, N.; Nakayama, K.; Nakayama, H.
Mol. Gen. Genet. 205, 298-304, 1986
A:Title: The recQ gene of Escherichia coli K-12: primary structure and evidence for SOS
A:Accession: J03137; MUID:87115164; PMID:3027506
A:Molecule type: DNA
A:Residues: 1-256, 'A', 258-610 <IRT>
A:Experimental source: strain K12
A:Note: It is uncertain whether residue 1, 3, 8 (all coded by GTG), or 9 (coded by TTG)
R:Rumezu, K.; Nakayama, K.; Nakayama, H.
Proc. Natl. Acad. Sci. U.S.A. 87, 5363-5367, 1990
A:Title: Escherichia coli recQ protein is a DNA helicase.
A:Reference number: A35776; MUID:90319113; PMID:2164680
A:Accession: A35776
A:Status: Preliminary
A:Molecule type: protein
A:Residues: 4-8 <UMES>
R:Daniels, D.L.; Plunkett III, G.; Burland, V.; Blattner, F.R.
Science 257, 771-778, 1992
A:Title: Analysis of the Escherichia coli genome: DNA sequence of the region from 84.5 to
A:Reference number: S30660; MUID:92358234; PMID:1379743
A:Accession: S30712
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 'V', 2-256, 'A', 258-610 <DNA>
A:Cross-references: EMBL:M87049
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1992
C:Genetics:
A:Gene: recQ
A:Map position: 85 min
A:Start codon: GTG
C:Function:
A:Description: involved in the recQ recombination pathway; its gene expression is under
A:Pathway: recQ protein; recQ helicase homology
C:Superfamily: ATP; DNA binding; hydrolase; nucleotide binding; P-loop; recQ recombination
C:Keywords: ATP; DNA binding; hydrolase; nucleotide binding; P-loop
F:49-56/Region: nucleotide-binding motif A (P-loop)
F:144-149/Region: nucleotide-binding motif B
F:148-151/Region: DEAH motif
F:365-402/Domain: recQ helicase homology <RHH>

Query Match          9.8%; Score 631.5; DB 1; Length 610;
Best Local Similarity 37.6%; Pred. No. 4.1e-26;
Matches 164; Conservative 63; Mismatches 134; Indels 75; Gaps 15;

QY      474  EQLGHQAFRPGQEAVMRIISGISTLVLPFGAGSKSCYOLPALLYSRSPCLTVVSPL 533
DB      21  ETFGYQGRPGQEEIITVLSGRDCLVVMPPGGKSLCTQIPALLNG-----LTVVSPL 76
QY      534  .LSLMDDOVSGLP-CLRAACHTSGMTRKQRESVYLOKTRAAOVHVLMTPEALVGAGLPP 592
DB      77  ISLMKDQVQLOANGVAAACLNSTOTREQOLEVMTGCRGTQIRLLVYLAPEELMDNPLEH 136
QY      593  AAQLPPIPAFAICIDEAHCLSSWSHNRPFCYLAVCVLAEEBNCVHOFGLGTATARTASDV 652
DB      137  LAHNHPVLLA-VDBAHCTISQWGHDFREYALAG-LKQRPPTLPFMAITATADDTTROI 194
QY      653  AQHLAAVEEPPHAPAPVPTMLHLSVMDRDTQALLTLQSKRFLWDSI-----703
DB      195  VRLIG-----LNPPL-----IQIS-SFDRNIRYMLM-----EKFKPLDQLMRYVQORG 238
QY      704  ---IYCNREDTTRIALALFTCLHAAWPGSGRAPEKTTAEAYHAQKCSRRRRVQRAF 760
DB      239  KSGIITCNRSKAKVNDTARL-----QSKGISAAAHAGIENNVRAVDVEKF 284
QY      761  MGGOLRVVVAAGMGIDPRDVAVYHLGLPSPFESYVQAVGAGRGGQPAHCHLFL-- 818
DB      285  QRDDIQIVATVAAGMGINKNVAFVVFHFDIPRIESTYQETGAGAGDGLPAEAMLTYPD 344
QY      819  -----QPGQEDLRELRRH---VHADSTDLFAVKRLV-----QRFVPACTC- 855

```

Db 345 ADMANLRCLREKPOQ-LQDIERHKLNMGAFAPAQTCRRLVLTLYFEGEGQBPQGNCD 403
 QY 856 TCTRPPEQEGAVGE 871
 Db 404 ICLDPPKQYDGSSTDAQ 419

RESULT 3

F86069

ATP-dependent DNA helicase [imported] - Escherichia coli (strain O157:H7, substrain EDL95)

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #revision 16-Feb-2001 #ext_change 09-Jul-2004

C:Accession: F86069

R:Petra, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.U.; Mayhew

Iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: AB5480; MUID:21074935; PMID:11206551

A:Accession: F86069

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-611 <STO>

A:Cross-References: UNIPROT:O8X8N1; UNIPROT:O8FBM6; GB:AB005174; NID:g12518696; PIDN:AAQ

A:Experimental source: strain O157:H7, substrain EDL953

C:Genetics:

A:Gene: recQ

C:Superfamily: recQ protein; recQ helicase homology

Query Match 9.8%; Score 631.5; DB 2; Length 611;
 Best Local Similarity 37.6%; Pred. No. 4,1e-26;
 Matches 164; Conservative 63; Mismatches 134; Indels 75; Gaps 15;

QY 474 EOLGHQAPRPGGERAVMRILSGISTLLVPTGAGKSLCYQLPALYSRRSPCLTLVSP 533
 Db 21 ETFGIGQFRPGQBEIITDVLSGRDCLVMPGTGGKSLCYQIPALLNG---LTVVSP 76
 QY 534 ISLMDQVSGLP-CLKAACIHSGMTKQRESVLOKIRAAQVHVMLTPEALVAGAGLPP 592
 Db 77 ISLMDQVQDQANGVAAACLNSTQTRQGLEMTGCRGQIRLTYIAPERLMDNPLEH 136
 QY 593 AAQLPPVAFACIDEAHCLISOMSHNRPCTLRVCKVLREMGVHCFGLTATRTTASDV 652
 Db 137 LAHNPVLLA-VDEAHCLISOMGHDRPEYALGQ-LRQRPPTLPFALATATDTRTODI 194
 QY 653 AQHLVAEERPDHGPAPVTNLHLSVMDRDTQALLTLQGRFONDSI----- 703
 Db 195 VRLG-----LNDPL-----IQIS-SFDRPPIRMLM---EKFRPDLQMRVYQEQRG 238
 QY 704 ---IYCNREDTERIALLRCTLHAAMVPGSGGAPKTTAEVYHAGMCSRRRRVQRAF 760
 Db 239 KSGIIVCNSRAKVEDTAARL-----QSKGISAAVYHAGLENNVRADVQERF 284
 QY 761 MGOGLRVVATVAFGMDRPRVAVLHGLPSPFSYQAVGRAGDOPAHCHLFL-- 818
 Db 285 QRDDLIQVATVAFGMDRPRVAVHEDIRNIESYQETGRAGRDGLPFAAMLFDYP 344
 QY 819 -----QPGEDLRELRRH---VHADSTDFLAVKRLV-----QVFPATC- 855
 Db 345 ADMANLRCLREKPOQ-LQDIERHKLNMGAFAPAQTCRRLVLTLYFEGEGQBPQGNCD 403
 QY 856 TCTRPPEQEGAVGE 871
 Db 404 ICLDPPKQYDGSSTDAQ 419

RESULT 4

H91222

ATP-dependent DNA helicase [imported] - Escherichia coli (strain O157:H7, substrain R1MD

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #ext_change 09-Jul-2004

C:Accession: H91222

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: H91222

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-611 <HAY>

A:Cross-References: UNIPROT:O8X8N1; UNIPROT:O8FBM6; GB:BA000007; PIDN:BA38175.1; PID:g1

A:Experimental source: strain O157:H7, substrain R1MD 0509952

C:Genetics:

A:Gene: ECG4752

C:Superfamily: recQ protein; recQ helicase homology

Query Match 9.8%; Score 631.5; DB 2; Length 611;
 Best Local Similarity 37.6%; Pred. No. 4,1e-26;
 Matches 164; Conservative 63; Mismatches 134; Indels 75; Gaps 15;

QY 474 EOLGHQAPRPGGERAVMRILSGISTLLVPTGAGKSLCYQLPALYSRRSPCLTLVSP 533
 Db 21 ETFGIGQFRPGQBEIITDVLSGRDCLVMPGTGGKSLCYQIPALLNG---LTVVSP 76
 QY 534 ISLMDQVSGLP-CLKAACIHSGMTKQRESVLOKIRAAQVHVMLTPEALVAGAGLPP 592
 Db 77 ISLMDQVQDQANGVAAACLNSTQTRQGLEMTGCRGQIRLTYIAPERLMDNPLEH 136
 QY 593 AAQLPPVAFACIDEAHCLISOMSHNRPCTLRVCKVLREMGVHCFGLTATRTTASDV 652
 Db 137 LAHNPVLLA-VDEAHCLISOMGHDRPEYALGQ-LRQRPPTLPFALATATDTRTODI 194
 QY 653 AQHLVAEERPDHGPAPVTNLHLSVMDRDTQALLTLQGRFONDSI----- 703
 Db 195 VRLG-----LNDPL-----IQIS-SFDRPPIRMLM---EKFRPDLQMRVYQEQRG 238
 QY 704 ---IYCNREDTERIALLRCTLHAAMVPGSGGAPKTTAEVYHAGMCSRRRRVQRAF 760
 Db 239 KSGIIVCNSRAKVEDTAARL-----QSKGISAAVYHAGLENNVRADVQERF 284
 QY 761 MGOGLRVVATVAFGMDRPRVAVLHGLPSPFSYQAVGRAGDOPAHCHLFL-- 818
 Db 285 QRDDLIQVATVAFGMDRPRVAVHEDIRNIESYQETGRAGRDGLPFAAMLFDYP 344
 QY 819 -----QPGEDLRELRRH---VHADSTDFLAVKRLV-----QVFPATC- 855
 Db 345 ADMANLRCLREKPOQ-LQDIERHKLNMGAFAPAQTCRRLVLTLYFEGEGQBPQGNCD 403
 QY 856 TCTRPPEQEGAVGE 871
 Db 404 ICLDPPKQYDGSSTDAQ 419

RESULT 5

AH0917

ATP-dependent DNA helicase (EC 3.6.1.-) [imported] - Salmonella enterica subsp. enterica

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #ext_change 18-Nov-2002

C:Accession: AH0917

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.,

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AH0917

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-609 <PAR>

A:Cross-References: GB:AL513382; PIDN:CAD07934.1; PID:g16504479; GSPDB:GN00176

C:Genetics:

A:Gene: recQ

C:Superfamily: recQ protein; recQ helicase homology


```
Db      246 ORGQ---CGIIYCGSRKKYEMLTU---C-----GNHR-----AASYHGMADDERA 288
Qy      755 RYQRAFMGQQLRVVAVTATVAFGWLGRPDVAVLHGLPSPFSSYQAVGRAGDGPANC 814
Db      289 WVOEAFORDDLOIVAVTAVAFGWLKPNVRPVVHPDIPNIESYYOETGRAGRDGLPAEA 348
Qy      815 HLFLOPQGEDLRELRRHVADSTDFLAVRLVQVRVPACTCTPRPSPQEGAVGGERPV 874
Db      349 MMLYTP--ADMMWMLRML-----DEKPD 369
Qy      875 PKYPOEAEQLSHQAAPGRRVC-----MGHERALPIQTLVQALDMP-----EBAI 920
Db      370 GAQKQVESHKLTAMGAFAPACRCRQVLLNMFGEYRDKRCGNCIDICLDPKPFDAEABAR 429
Qy      921 ETLIC 925
Db      430 KALSC 434
```

RESULT 8

```
H97365
DNA helicase Xf1381 [Imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C.Species: Agrobacterium tumefaciens
C.Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C.Accession: H97365
R.Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollem, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A.Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A.Reference number: A97359; PMID:21608551; PMID:11743194
A.Accession: H97365
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-602 <KUR>
A.Cross-references: UNIPROT:Q8UJ77; GB:AE007869; PIDN:AAK85881.1; PID:g15154924; GSPDB:G
A.Genetic:
A.Gene: AGR C 92
A.Map position: circular chromosome
C.Superfamily: recQ protein; recQ helicase homology
```

```
Query Match      8.7%; Score 560.5; DB 2; Length 602;
Best Local Similarity 34.6%; Pred. No. 2,2e-22;
Matches 151; Conservative 73; Mismatches 152; Indels 61; Gaps 17;

Qy      466 PAEFOALBQLGHAFRPQGERAVNRILSGISTLVLPPTGAGKSLCYQLPALYSRRSPC 525
Db      5 PLNLIKAV--YGYDAFRGRQGERIIQHVAGNNAFVLMPTGGKSLCYQIPAL--AREG-- 58
Qy      526 LTLVVSPLSLMDDDVSGLRPC-LKAACIHSGMTKQRESVQKIRAAQVHMLMLPREAL 584
Db      59 MGLVVSPLIALMWDVVALRQAGVRAEALNSDLSPEERTLRDVRAGVNDILYAAPELT 118
Qy      585 VAGAGLPAAQLPVAFACTIDEAHCLSQWNSHFRPCYLRCVKLBERMVGHFGLTLATA 644
Db      119 LKPDVL-DLQSIDSLAVDAHCLSQWNSHFRPCYLRCVKLBERMVGHFGLTLATA 176
Qy      645 TERTSDVAQHLVAEEDPLHGPAPVPTNLHLSVSMRDTQALLTLQKRFQ--LD 701
Db      177 DEPTRAEILGHIAL-DEADAFIAGFDRPNIRYAI-MEKDNPTQL-----KFLNGRENE 229
Qy      702 STIITCNREDTERIALALRTCLHAAMVPGSGGRAPKTTAEVYHAGMCRERRRQORAFM 761
Db      230 SGIVYCLSRKRYEETAAWL-----EBGR-----DALPYHAGMDRAAEENQTRFO 275
Qy      762 OGQLRVAVATVAFGWLGRPDVAVLHGLPSPFSSYQAVGRAGDGPANC----- 814
Db      276 HGEAVIYATVAFGWLGRPDVAVLHGLPSPFSSYQAVGRAGDGPANC----- 814
Qy      815 -----HLFLOPQGEDLRELRRHVADSTDFL-----AVKRLV-----QVFPACTC- 855
Db      336 DIALNRRTIE--SDAADQKRYMEROKLDALGLAETAGCRQVLLSYFGDRCEPCGNC 393
```

```
Qy      856 TCTRPSPQEGAVGGER 872
Db      394 TCSSPPDLFEGALIAOK 410
```

RESULT 9

```
AH2583
ATP-dependent DNA helicase recQ [Imported] - Agrobacterium tumefaciens (strain C58, Dupo
C.Species: Agrobacterium tumefaciens
C.Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C.Accession: AH2583
R.Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A.Authors: Yoo, H.; Teo, Y.; Biddle, P.; Jung, M.; Krepsan, W.; Perry, M.; Gordon-Kamm,
ster, B.W.
A.Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A.Reference number: AB2577; PMID:21608550; PMID:11743193
A.Accession: AH2583
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-602 <KUR>
A.Cross-references: UNIPROT:Q8UJ77; GB:AE008688; PIDN:AAL41086.1; PID:g17738376; GSPDB:G
A.Experimental source: strain C58 (Dupont)
A.Genetic:
A.Gene: recQ
A.Map position: circular chromosome
C.Superfamily: recQ protein; recQ helicase homology
```

```
Query Match      8.7%; Score 560.5; DB 2; Length 602;
Best Local Similarity 34.6%; Pred. No. 2,2e-22;
Matches 151; Conservative 73; Mismatches 152; Indels 61; Gaps 17;

Qy      466 PAEFOALBQLGHAFRPQGERAVNRILSGISTLVLPPTGAGKSLCYQLPALYSRRSPC 525
Db      5 PLNLIKAV--YGYDAFRGRQGERIIQHVAGNNAFVLMPTGGKSLCYQIPAL--AREG-- 58
Qy      526 LTLVVSPLSLMDDDVSGLRPC-LKAACIHSGMTKQRESVQKIRAAQVHMLMLPREAL 584
Db      59 MGLVVSPLIALMWDVVALRQAGVRAEALNSDLSPEERTLRDVRAGVNDILYAAPELT 118
Qy      585 VAGAGLPAAQLPVAFACTIDEAHCLSQWNSHFRPCYLRCVKLBERMVGHFGLTLATA 644
Db      119 LKPDVL-DLQSIDSLAVDAHCLSQWNSHFRPCYLRCVKLBERMVGHFGLTLATA 176
Qy      645 TERTSDVAQHLVAEEDPLHGPAPVPTNLHLSVSMRDTQALLTLQKRFQ--LD 701
Db      177 DEPTRAEILGHIAL-DEADAFIAGFDRPNIRYAI-MEKDNPTQL-----KFLNGRENE 229
Qy      702 STIITCNREDTERIALALRTCLHAAMVPGSGGRAPKTTAEVYHAGMCRERRRQORAFM 761
Db      230 SGIVYCLSRKRYEETAAWL-----EBGR-----DALPYHAGMDRAAEENQTRFO 275
Qy      762 OGQLRVAVATVAFGWLGRPDVAVLHGLPSPFSSYQAVGRAGDGPANC----- 814
Db      276 HGEAVIYATVAFGWLGRPDVAVLHGLPSPFSSYQAVGRAGDGPANC----- 814
Qy      815 -----HLFLOPQGEDLRELRRHVADSTDFL-----AVKRLV-----QVFPACTC- 855
Db      336 DIALNRRTIE--SDAADQKRYMEROKLDALGLAETAGCRQVLLSYFGDRCEPCGNC 393
```

RESULT 10

```
AD1419
ATP-dependent DNA helicase homolog lmo2757 [Imported] - Listeria monocytogenes (strain
C.Species: Listeria monocytogenes
C.Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C.Accession: AD1419
```

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fehl, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Krett, U.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitouram, A.; Ma
ok, C.; Schuener, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of *Listeria* species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AD1419
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-590 <GLA>
A;Cross-references: UNIPROT:O8Y3S4; GB:NC_003210; PIDD:CAD000970.1; PIDD:g16412257; GSPDB
A;Experimental source: strain ESD-e
C;Genetics:
A;Gene: lmo2757

[illegible]

RESULT 11
 AE1832
 ATP-dependent DNA helicase [imported] - Nostoc sp. (strain PCC 7120)
 C|Species: Nostoc sp. PCC 7120
 A|Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C|Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
 C|Accession: AE1832
 R|Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Matanabe, A.; Iriuguchi,
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A|Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A|Reference number: AB1807; MUID:21595285; PMID:11759840
 A|Accession: AE1832
 A|Status: Preliminary
 A|Molecule type: DNA
 A|Residues: 1-718 <KUR>
 A|Cross-references: UNIPROT_Q8Z093; GB:BA000019; PDB:BAW7729.1; PDB:gl7135183; GSPDB:C
 A|Experimental source: strain PCC 7120
 C|Genetics:
 A|Gene: alr0205
 C|Superfamily: recQ protein; recQ helicase homology

[illegible]

```

RESULT 12
AEI1794
APP-dependent DNA helicases homolog lin2900 [imported] - Listeria innocua (strain Clij11
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AEI1794
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feihl, H
D.: Jones, L.M.; Karet, U.
Science 294, 849-852, 2001
A:Authors: Krett, U.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madeno, E.; Maitournam, A.; Me
Ok, C.; Schlueter, T.; Simoes, N.; Tlertsz, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AEI1794
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-590 <GLA>
A:Cross-references: UNIPROT:Q926Z0; GB:AL592022; PIDN:CAC98125.1; PID:G16415441; GSPDB:C
A:Experimental source: strain Clij11262
C:Genetics:
A:Gene: lin2900
C:Superfamily: recQ protein; recQ helicase homology

Query Match      8.5%, Score 548; DB 2; Length 590;
Best local Similarity 36.5%, Pred. No. 1e-21;
Matches 132; Conservative 65; Mismatches 127; Indels 38; Gaps 9;

QY      474 BOLGQAERFQGEBAVNRKILSGISTLLVLPFGAGKSLCYQLPALVYRRSPCLTLVVSPL 533
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      11 QNFGQPRDQDIVISKLSAGEDTLAIMPTGGKSLCYQLPALFPDQ---LTIIVSPL 66

QY      534 LSLMDQVYSGL-PCLKRAACIHSGMTKQKRSVYQKIRAAQVHYMLTLPALVAGGLRP 592
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      67 LSKMDQDALVSDGIAATFNLSTLRREIDIRLDAAFSGGLKXLYLAPEKIEFTPGQRL 126

```

QY 593 AAOLPPVAFACIDEAHCLCSOMSHNPRCYLVRCKVLEBNGVHCFLGTATATRTASPDY 652
 DB 127 IQGVPSLSFA-IDENACISOMGHDFRPSLISLSDSKTRRLVLTALITATQAASDDI 185
 QY 653 AOHLLAAVEEPDLHGPAVPPT-----NLHLSVSMRDRDQALLTLLOGKRFONLDSIIIVC 707
 DB 186 CRLTKI--KPN-----SVIKTGFERNLAFQVNVGQDKYLDYLRK-----NSTSEGIITVA 236
 QY 708 NNRREDTERIALRLTCLHAAMPVSGSGRAKTTAE--YHAGCSERRRVRQAPMOGOL 765
 DB 237 STRKEERHISFL-----KKGESEGMHGGMTDARXDMOKEFLYDDI 280
 QY 766 RVVAVTAVAGMGDLRDPVAVLHGLPSPFESYVQAVGRAGRDGPAHCHLFLPOGEDI 825
 DB 281 RIVVATNAFGMGKINSNVAFVHYNIPIRTEATYQAGRAGRDGVPSDCLLFPDSKRI 340
 QY 826 RE 827
 DB 341 QQ 342

RESULT 13

ES3226
 ATP-dependent DNA helicase RecQ PA3344 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
 C:Accession: E83226
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardi, K.; Lam,
 .J.; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 gen
 A:Reference number: A82950; PMID:20437337; PMID:10984043
 A:Accession: E83226
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-712 <STO>
 A:Cross-references: UNIPROT:Q9HVO1; GB:AE004757; GB:AE004091; NID:G9949476; PIDD:ANG0673
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: recQ; PA3344
 C:Superfamily: recQ protein; recQ helicase homology

Query Match 8.5%; Score 545; DB 2; Length 712;
 Best Local Similarity 28.5%; Pred. No. 1,8e-21;
 Matches 221; Conservative 102; Mismatches 294; Indels 158; Gaps 32;

QY 471 QALBOL-----GHOAFRPGQERAVMRILSGISTLLVLTGAGSKSLCYQLPALYSRRSPCL 526
 DB 4 QALRLKGVFGYDARFRGNQARIIEVAEGDALVIMPTGGKSLCFQVALLAREG-----L 59
 QY 527 TLVSPPLSLMDQVS-----GLPCLKACIHSGMTKRRQBSVLQKIRAAOVHVLMLTP 581
 DB 60 TVVSPPLSLALMEDQVATDELGV-----AVALNSTLNPQORDIAERLGRGRIKLLYLAP 115
 QY 582 BALVAGGIPPAACLPVAFACIDEAHCLCSOMSHNPRCYLVRCKVLEBNGVHCFLGT 641
 DB 116 ERLVPRMLAFQRLPVGILFA-IDENACISOMGHDFRPSLISLSDSKTRRLVLTALITATQAASDDI 173
 QY 642 ATATRTASDVAAQHLLAAVEEPDLHGPAVPPTMHLVSVMRDRDQALLTLLOGKRFONL 701
 DB 174 ATADKRTREEMTQRHLQNAEQFLSFDPRNIFVYIVREGQRKQ-LIGFLSERG--D 229
 QY 702 SIITICNRREDTERIALRLTCLHAAMPVSGSGRAKTTAEYHAGCSERRRVRQAPM 761
 DB 230 AGIVVCLSRKKEVAEFL-----GNQ-----FPALPYHAGLSNELAHNQKRL 275
 QY 762 OGQLVWVAVTAVAGMGDLRDPVAVLHGLPSPFESYVQAVGRAGRDGPAHCHLFLPOG 821
 DB 276 NEEGLIMATIAFGMGKINSNVAFVHYNIPIRTEATYQAGRAGRDGVPSDCLLFPDSKRI 340
 QY 822 GEDLLERRHVADSTDLAVKRLVQVRPACTCTCTTRPSEOGAVG---SERVP--- 875

DB 336 ---DVILLRQMQSSBGD-ERHKRVERHKLEAMLALCEETRCRQALLAYFDEMPQCGH 392
 QY 876 -----KTPPOEABOLSHQAP---GPRVCKGERALPIQTVQALDMPBEAETLLC 925
 DB 393 CINCVDGVEWTWATESARQALSAIYSGQRYGVGH-----LVDTLLGRETEKIRSL-- 443
 QY 926 YLEHNN-----HW-----LELLATYTHCRINCQGPQALQALNRCPL-- 966
 DB 444 ---GHQHLAVFGIGKGRGEDEWRTFLRQLVAAGLADVDLDGRG---LRLTEACRPLR 496
 QY 967 ---AVCLAQQLBEDPQSGSS---SVEFDMVLKDSMGWE-LASVRLACQLQMDPEPT 1018
 DB 497 GEVRLERLRLDLRQAKSSSGSAGASQVYSEBEMMALRLKLAE---EHSVPP 553
 QY 1019 GVRRTGVLVPESELAFLHRSFGDLTAEEKDQICPLYGVQARERQALRLRTFOAP- 1077
 DB 554 YVIFPDATLEW--LRSPRSLDMAQ-----VSGVGR-----KLERYQALFL 595
 QY 1078 -----HSVAFPSG-GPCLRQOEERSTRKDLLGRYFEEBEGQEPG 1118
 DB 596 DVLTSPPAAPAPPODLRHELASLACAGTPAQIARQALNCSKENVAMLAIAQGVSL 655
 QY 1119 MEDAQPPEPGQARLDQMEDQVRCIRQFL---SLRP---EKTSSRAVARIFH 1165
 DB 656 EQALDLPBELLGIDQ-----AFLEBDGELPVAALBERGKRVPSGLH 700

RESULT 14

A69691
 ATP-dependent DNA helicase homolog - Bacillus subtilis
 N:Alternate names: ATP-dependent DNA helicase recQ
 C:Species: Bacillus subtilis
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
 C:Accession: A69691
 R:Kunet, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Evington, J.; Fabre, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; PMID:98044033; PMID:9384377
 A:Accession: A69691
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-496 <KUN>
 A:Cross-references: UNIPROT:P50729; GB:Z99115; GB:Z99116; GB:AL009126; NID:52634723; PID
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: recQ
 C:Keywords: ATP; nucleotide binding; P-loop
 F:38-45/Region: nucleotide-binding motif A (P-loop)
 F:132-137/Region: nucleotide-binding motif B
 F:136-139/Region: DNA motif

Query Match 8.4%; Score 542; DB 2; Length 496;
 Best Local Similarity 36.8%; Pred. No. 1.7e-21;
 Matches 137; Conservative 68; Mismatches 117; Indels 50; Gaps 12;

QY 471 QALBOL-----LGHQAFRPGQERAVMRILSGISTLLVLTGAGSKSLCYQLPALYSRRSPCLTV 529
 DB 6 QTLVQFPFTSKKQDQIISLISGKDTIALPLPGGKSLCYQLPGYMLDG---MVLTI 61
 QY 530 VSPPLSLMDQVSGLPCC--LKAACIHSGMTKRRQBSVLQKIRAAOVHVLMLTPALVGA 587

Db 62 VSPILSLMEDVOQLKARGEKRAALNSMLNRQERQVLEHIIH--RYKFLYLSPEALOSP 119
 QY 588 GGLPPAALPVPVAFCTIDEAHCLISQWNSHNPFCYLRVCKVLEREMGVHCFGLTATATRR 647
 Db 120 YLEKTLKSV-PLSLFVIDEABHCTISEMGHDFRDPYSLQ-LRKKGHPVPLTLTATATKE 177
 QY 648 TASDV-----AQLVAVAEPPDLHGPAVPPTNHLVSMSDRDTPDQALLTLQGRFQ 698
 Db 178 TLQDWMNLLELGHAVRHLSNVNR-----NIALRVENAADTAEKIDRVIQ--LVE 225
 QY 699 NLDSI-IYCNRRRETERIAALLRTCLHAANVPGGGAPKTTAA--YHAGMCRERRR 755
 Db 226 NLQGPYIYCPFRKAKELAGEIKS-----KTSRADPFGHGLBSGDRIL 270
 QY 756 VQARMOGLRVVAVTVAFGMDLDRPDYRAVLHLGLPSFESYQAVAGRAGRDQPAHCH 815
 Db 271 IQQDFINQDLVDYICCTNAFGMGVDKPDIRYVHFHPLTAERAFMOEIGRAGDKRPSVI 330
 QY 816 LFLDPQSGDLRE 827
 Db 331 LLRARGDFELQE 342

RESULT 15

S62467

ATP-dependent DNA helicase hus2 - fission yeast (Schizosaccharomyces pombe)

C.Species: Schizosaccharomyces pombe

C.Date: 16-May-1996 #sequence revision 13-Mar-1997 #text_change 09-Jul-2004

C.Accession: S62467; T38578; T43265

R.Badcock, K.; Churcher, C.M.

submitted to the EMBL Data Library, October 1995

A.Reference number: S62445

A.Accession: S62467

A.Status: preliminary

A.Molecule type: DNA

A.Residues: 1-1328 <BAD>

A.Cross-references: UNIPROT:Q09811; EMBL:254354; NID:g1019398; PIDN:CAA91177.1; PID:g101

R.Badcock, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.

submitted to the EMBL Data Library, October 1995

A.Reference number: Z21745

A.Accession: T38578

A.Status: preliminary; translated from GB/EMBL/DBJ

A.Molecule type: DNA

A.Residues: 1-1328 <BA2>

A.Cross-references: EMBL:254354; PIDN:CAA91177.1; GSPDB:GN00066; SPDB:SPAC2G11.12

A.Experimental source: strain 972h; cosmid c2G11

R.Stewart, E.; Chapman, C.R.; Al-Khodairy, F.; Carr, A.M.; Enoch, T.

EMBO J. 16, 2682-2692, 1997

A.Title: Rqh1+, a fission yeast gene related to the Bloom's and Werner's syndrome genes,

A.Reference number: Z07997; MUID:97327548; PMID:9184215

A.Accession: T43265

A.Status: preliminary; translated from GB/EMBL/DBJ

A.Molecule type: DNA

A.Residues: 1-1328 <STB>

A.Cross-references: EMBL:Y09426; NID:g1684753; PIDN:CAA70577.1; PID:g1684754

C.Genetics:

A:Gene: hus2; SPAC2G11.12

A:Map position: 1L

C:Superfamily: Bloom's syndrome helicase; recQ helicase homology

C:Keywords: ATP; nucleotide binding; P-loop

F:541-548/Region: nucleotide-binding motif A (P-loop)

F:647-652/Region: nucleotide-binding motif B

F:651-654/Region: DEAH motif

F:875-914/Domain: recQ helicase homology <RHH>

Query Match 8.2%; Score 525.5; DB 2; Length 1328;
 Best Local Similarity 37.1%; Pred. No. 3.8e-20;
 Matches 143; Conservative 49; Mismatches 154; Indels 39; Gaps 11;

QY 443 VPSLPTVLYSLPSGGLAETPAVFPALQLGH-QAFRPGQERAVNRITLIGISTLLV 501
 Db 491 MPSLDD---PMLSYPMW-----KVLGCLKHKFFHLKGRKQQLKALNGITLISGKDVFTL 540

QY 502 LPTGAGKSLCYQLPALVYSRRSPCLTLVVSPLSLMDQVSGLP---PCLKAACIHSQM 557
 Db 541 MPTGGKSLCYQLPAVIEGASRGVTLVISPILSIMQQLHLRLKLNIPSLPISGEQPAD 600
 QY 558 TRQRESVLOKIRAAQVNVMLTPALVAGG---LPPAQLPVPVAFACIDEAHCISQW 613
 Db 601 ERQVTSFLM-AKVNLVTLVYTPPEGLASNGAITRVLSLYERKLLARIVIDEACVSHW 659
 QY 614 SHNPRCYLRVCKVLEREMGVHCFGLTATATRTASDVAQHLVAEPPDLHGPAVPPTN 673
 Db 660 GHDFRPDYKQ-LGLRDYQGIPEFVALTRTANETVKDIIWTLRMENCLLEKSSFNRR-N 717
 QY 674 LHLVSMDRDTQALLTLQGRFPQNLDSIIYCNRRRETERIAALLRTCLH-AAWPGS 732
 Db 718 LFYEIKPKDLYTELRYFISNGHLH--ESGIYCLSRFSCQVAKLNDYGLKAM----- 771
 QY 733 GGRAPKTAELVHAGMCRERRRRVQRAVMOQLRVVAVTVAFGMDLDRPDYRAVLHLGLP 792
 Db 772 -----HYHAGLEKVERQRIQNEWQSGSYKTIIVATIAFGMGVDKGDVRFVTHSFP 821
 QY 793 PSFESYQAVAGRAGRDQPAHCHLP 817
 Db 822 KSLBGIYQETGRAGRDGKPAHCIMF 846

Search completed: March 19, 2005, 01:09:44
 Job time : 54 secs